THE USE AND INTERPRETATION OF EDUCATIONAL TESTS

HARRY A. GREENE
ALBERT N. JORGENSEN

LONGMANS' EDUCATION SERIES

LIBRARY

OHIO NORTHERN UNIVERSITY

REGULATIONS

1. Books may be borrowed for a period of two weeks.
2. At the expiration of two weeks books may be re-let if returned to the librarian in good condition.
3. A rental of two cents a day will be charged on each book held over two weeks.

Class No. 3716 Accession No. 21349

erick Memorial Library
o Northern University

THE USE AND INTERPRETATION OF EDUCATIONAL TESTS

Longmans' Education Series

Edited by

FREDERIC B. KNIGHT

Professor of Education and Psychology
in the University of Iowa

The Psychology of Abnormal People. *By* JOHN J. B. MORGAN, *Associate Professor of Psychology, Northwestern University.*

Work-Book in Abnormal Psychology [Forms A and B]. *By* JOHN J. B. MORGAN, *Associate Professor of Psychology, Northwestern University.*

Work-Book in Educational Measurements [Forms A and B]. *By* HARRY A. GREENE, *Associate Professor of Education, University of Iowa.*

The Use and Interpretation of Educational Tests. *By* HARRY A. GREENE, and ALBERT N. JORGENSEN, *Professor of Education, Michigan State Normal College, Ypsilanti.*

Study Manual in Elementary Statistics. *By* EVERETT F. LINDQUIST, *Assistant Professor of Education, University of Iowa,* and GEORGE D. STODDARD, *Professor of Psychology and Director of the Iowa Child Welfare Research Station, University of Iowa.*

Objectives and Procedures in Civic Education. *By* CHARLES C. PETERS, *Professor of Education, Pennsylvania State College.*

The Teaching of High School English. *By* VIRGINIA J. CRAIG, *Head of English Department, Southwestern Missouri State Teachers College.*

Making a High School Schedule of Recitations. *By* R. C. PUCKETT, *Director of High Schools, Toledo, Ohio.*

Introducing Education. *By* J. HERBERT BLACKHURST, *Professor of Education, Drake University*

IN PREPARATION

Mental Hygiene. *By* JOHN M. DORSEY, *Assistant Director of the Psychopathic Hospital, University of Michigan.*

Psychology of the High School Subjects. *By* H. H. REMMERS, *Professor of Education, Purdue University.*

Neuro-Physiology. *By* LEE EDWARD TRAVIS, *Professor of Speech Pathology, and Director of the Psychological and Speech Clinic, University of Iowa,* WAID WRIGHT TUTTLE, *Assistant Professor of Physiology, University of Iowa,* and FREDERIC B. KNIGHT.

Longmans' Education Series

THE USE AND INTERPRETATION OF EDUCATIONAL TESTS

A HANDBOOK FOR THE CLASSROOM TEACHER

BY

HARRY A. GREENE, PH.D.

ASSOCIATE PROFESSOR OF EDUCATION, AND
DIRECTOR OF BUREAU OF EDUCATIONAL RESEARCH AND SERVICE
UNIVERSITY OF IOWA

AND

ALBERT N. JORGENSEN, PH.D.

DIVISION OF EDUCATIONAL RESEARCH AND
ASSOCIATE PROFESSOR OF EDUCATIONAL ADMINISTRATION
MICHIGAN STATE NORMAL COLLEGE

LONGMANS, GREEN AND CO.
NEW YORK · LONDON · TORONTO

GREENE AND JORGENSEN
THE USE AND INTERPRETATION OF EDUCATIONAL TESTS

COPYRIGHT · 1929
BY LONGMANS, GREEN AND CO.

First Edition January 1929
Reprinted July 1929
February 1930, August 1930
October 1931

MADE IN THE UNITED STATES OF AMERICA

79

PREFACE

THIS book is designed to place in the hands of the classroom teacher and the student of education an introductory and elementary discussion of the essential principles of measurement in education. The material and the plan of organization are the results of a number of years of experience in teaching courses in educational measurements and in practical work in this field with the classroom teacher. It is not a part of its purpose to compete with or to repeat the numerous more technical discussions of test construction and interpretation now available. These are for the advanced statistician and research student. The present treatment is planned to provide a systematic handbook for any person who feels the need of a well-organized treatment of the fundamental ideas and tools of measurement in education. It is written from the practical point of view of the classroom teacher, and in language sufficiently non-technical to appeal to the beginning student in this field which to some seems highly technical. It is hoped that because of this direct method of attack, the practical point of view taken, and the concrete illustrations, it may appeal to the group for which it is intended. Only through the active enlistment of the interest of the classroom teacher and of the student who himself is to serve in that capacity will the fullest possibilities of measurement in education be realized.

The authors believe that the chief purpose served by standard tests is the improvement of classroom instruction through the more exact location, diagnosis, and remedy of individual pupil difficulties in learning. They believe, too, that the responsibility for the use of these instruments for the location,

21349

diagnosis, and remedy of pupil difficulties must be more and more assumed by the classroom teacher. This does not mean any less responsibility on the part of the supervisor. It means a clearer recognition of it on the part of the classroom teacher, for after all the use and interpretation of educational tests is a coöperative project. It is all directed to the same end, the goal of all good supervision, the improvement of the conditions under which teachers teach and children learn.

To accomplish these purposes, much more emphasis must be placed upon the training of the prospective teacher, as well as of the more experienced teacher, in the use of measuring devices and in the development and application of remedial materials following diagnosis. This book is an attempt to place in the hands of the student, the classroom teacher, and the supervisor some basic suggestions of ways in which educational measuring instruments may be of practical use in the teaching of children and some general rules for guidance in selecting, using, and interpreting educational tests as valuable aids in accomplishing this task. If, through the reading of this brief treatment of the field as limited above and the working of the practical problems in measurement given in the *Appendix,* the teacher or student shall be prompted to a more intensive examination of the possibilities of standard tests as teaching and supervisory instruments, the purpose of this book will be fulfilled.

The indebtedness of the authors to the many experienced teachers and supervisors who have contributed indirectly as well as directly to the formulation of much of the material incorporated herein is gratefully acknowledged. Special thanks are due to Miss Mabel Snedaker, Extension Supervisor, University of Iowa, and to Miss Maud McBroom, Supervisor of the University Elementary School, as well as to the teachers of this school for the suggestion of a number of remedial devices in the field of reading and language. The authors' special thanks are also due Dr. G. M. Ruch, Professor of Education, University of California, who read much of the

manuscript and made many helpful suggestions as to its organization. We are also grateful to the authors and publishers who gave permission for the reproduction of many selected portions of their work and publications. Specific acknowledgments of these are made throughout the text.

We wish also to acknowledge our very deep indebtedness to Professor Frederic B. Knight for his careful study of this manuscript and for his many valuable suggestions as to organization and content, and to Miss Neva Steele, Miss Esther Flynn, and Mr. Paul K. Frazer for their assistance in the final preparation of the manuscript.

EDITOR'S INTRODUCTION

Professor Greene and Dr. Jorgensen have had close contact with the problems of the improvement of classroom instruction as well as adequate technical training. Such experience qualifies them in a happy way to understand not only Educational Tests themselves but also the needs of teachers and supervisors. A union of practical experience and technical training possessed by the authors has produced a book on measurements which is sound in theory and effective in presentation.

Doubtless, the expert in measurement will not find this book particularly illuminating. In fact, for him it may in many respects be an elaboration of the obvious. It is not written for him. However, the classroom teacher and the busy supervisor, and those in training for teaching or supervision will find this book a carefully written fundamental text on the principles of measurement in education. Its main contribution to the growing literature on measurement is, I think, not so much in new points of view or advances in the technicalities of test construction but rather in its plainness of exposition and balance in treatment of many points which to some seem over technical. It should be thought of as a first book in measurement for those who do not at the time of studying it know very much, if anything, about measurement in education and its application to the problems of improving classroom instruction.

A careful examination of the book will support the opinion that this book is not only excellent in content, as such, but is an effective learning instrument for students. A present

hopeful tendency of collegiate teaching is to stress factors of presentation as well as problems of content. To further increase the teaching and learning aspects of a first course in measurement Professor Greene has prepared a "Work-Book in Educational Measurements." The present book together with the Work-Book supplies to the college teacher and his pupils a well organized total learning unit in the field of educational measurements. Not only is the pupil taught, but he is given sufficient practice in measurement, so that upon mastery of the books he can face the problems of measurement in actual school situations with considerable confidence in his preparation to solve such problems with success.

Teachers must be well trained in the use and interpretation of educational tests. This book is definitely directed to meet in a frank and reasonable way the need for such training. It is a basic text for class-room use in a course on educational tests and their uses. It is also well adapted for use in connection with both Study Center and Reading Circle work.

F. B. Knight.

CONTENTS

CHAPTER PAGE

I. Introduction 1

I. Background of the Growth of the Measurement Idea.

1. Educational Measurements Not New Idea. 2. Real Pioneers in Measurements. 3. Factors in Growth of Measurement Idea. 4. Rapid Growth of Interest in Tests. 5. Purpose of this Book.

II. The Meaning of Educational Tests 10

I. What Tests Are.

6. Measurement in Education. 7. Characteristics of Educational Tests. 8. What Tests Do. 9. What Tests Do Not Do.

II. Kind of Tests.

10. Tests and Scales. 11. Mental Tests. 12. Individual Mental Examinations. 13. Group Tests of Mental Ability. 14. Special Aptitude and Prognostic Tests. 15. Tests of Achievement. 16. Quality and Product Scales. 17. Survey vs. Diagnostic Tests.

III. Summary.

III. The Uses of Educational Tests 47

18. Tests in Relation to Classroom Instruction.

I. Uses of Achievement Tests.

19. For Purposes of Comparison. 20. For Measuring Efficiency of Instruction. 21. For Purposes of Class Diagnosis. 22. For Purposes of Individual Pupil Diagnosis. 23. For Pupil Gradation and Guidance.

II. Uses of Mental Tests.

24. A Warning. 25. For Class Diagnosis. 26. For Individual Diagnosis. 27. Inter-

CHAPTER PAGE

dependence of Achievement and Mental Test Results.

III. Summary.

IV. Using Standard Tests in the Classroom..... 67

28. Introduction.

I. Giving Tests.

29. The Teacher's Responsibility. 30. Scoring and Interpreting Tests.

II. When to Give Tests.

31. Periodical vs. Continuous Use of Tests.

III. The Basis for the Testing Program.

32. A Clear-cut Teaching Problem as Basis. 33. Steps in a Testing Program. 34. Illustrations of Problems.

IV. The Criteria for Selecting Tests.

35. Validity. 36. Reliability. 37. Objectivity. 38. Educational Significance. 39. Administrability. 40. Economy. 41. Adequacy of Test in Measuring Specific Functions.

V. General Rules for a Testing Program.

42. Preparation for Administering Tests. 43. Timing. 44. Scoring.

VI. Summary.

V. Constructing and Using Informal Objective Examinations in the Classroom..... 89

I. Characteristics of Informal Objective Examinations.

45. Introduction. 46. The New Examination. 47. Informal vs. Standardized Tests.

II. Advantages and Disadvantages of the Informal Objective Examination.

48. Advantages of the Informal Objective Examination. 49. Disadvantages of the Objective Examination.

III. Principles of Constructing and Using Objective Exercises.

50. Need for Specific Directions. 51. Number of Exercises to Use. 52. Selection of the Subject-matter Items to be Tested. 53. Clear State-

CHAPTER PAGE

ment of Exercises. 54. Difficulty of the Exercises. 55. Arrangement of the Exercises. 56. Arrangement of Answers for Convenience in Scoring. 57. Use of Various Types of Exercises.

IV. Practical Use of Results.

58. Use of Objective Test Scores as Basis for Class Grades.

V. Classification and Illustrations of Types of Objective Exercises.

59. Classification of Objective Tests. 60. Samples of Types of Objective Test Exercises. 61. Chance Factor in Objective Tests.

VI. Summary.

VI. Summarizing the Results of Testing..... 116

62. Introduction.

I. Classification and Tabulation of Test Scores.

63. Need for a Method of Grouping Data. 64. Steps Involved in Preparing a Frequency Table. 65. Determining the Range of Scores. 66. Determining the Number and Size of Class-Intervals. 67. Determining the Limits of the Class-Intervals. 68. Tabulating the Scores. 69. Summary of Steps in Making a Frequency Table.

II. Measures of Central Tendency.

70. Computing the Arithmetic Mean from Ungrouped Data. 71. Computing the Arithmetic Mean—Data grouped in Frequency Distribution. 72. Summary of Steps in Finding Arithmetic Mean. 73. Use of Arithmetic Mean. 74. The Median as a Measure of Central Tendency. 75. Computing a Counting Median or Mid-Measure. 76. Computing the Median from Frequency Distribution. 77. Use of the Median.

III. Measures of Variability.

78. Need for Measures of Variability. 79. Range as an Expression of Variability. 80. "Q" or Semi-inter-quartile range as a Measure of Variability. 81. Computing the Quartile Deviation, "Q". 82. Further Practice in Find-

CHAPTER PAGE

ing "Q." 83. Standard Deviation (Sigma) as a Measure of Variability. 84. Computing the Standard Deviation. 85. Computing the Standard Deviation from Data in Frequency Distribution. 86. Additional Practice in Finding Sigma.

IV. The Relationship of Test Scores.

87. Need for Measures of Relationship. 88. Meaning of Correlation. 89. Tabulating Test Scores in Double Entry or Correlation Tables. 90. Steps in the Computation of the Pearson Product-Moment Coefficient of Correlation. 91. Illustration of the Computation of the Pearson Correlation Coefficient. 92. Meaning of the Correlation Coefficient. 93. Additional Experience with Correlations.

V. Assignment of Relative and Absolute Ranks.

94. Need for Method of Ranking. 95. Assignment of Relative Ranks. 96. Assignment of Absolute Ranks. 97. Computing Percentiles (Deciles) in a Frequency Distribution. 98. A Short Method of Computing Percentile Scores. 99. Computation of the Percentile Score Corresponding to a Given Test Score in a Frequency Table.

VI. Summary.

VII. Interpreting the Results of Testing 176

I. Test Norms and Standards.

100. Norms as Distinctive Features of Standardized Tests. 101. Norms vs. Standards. 102. Uses of Test Norms and Standards. 103. Kinds of Norms. 104. Derivation of Norms. 105. Meaning of Test Norms. 106. Grade Norms. 107. Age Norms. 108. Interpretation of Grade and Age Norms.

II. Types of Derived Scores.

109. Function of Derived Scores. 110. Grade Norms as Derived Scores—Type 1. 111. Age Scores as Derived Scores—Type 2. 112. Percentile Ranks as Derived Scores—Type 3. 113. Quotients as Derived Scores—Type 4.

III. Summary.

CHAPTER PAGE

VIII. The Place of Tests in Diagnosis and Remedial Teaching 207

I. The Function of Exact Diagnosis.

114. Importance of the Diagnostic Interpretation of Test Results. 115. Exact Diagnosis the Basis for Remedial Work. 116. Diagnosis as the Basis for Preventive Work. 117. General Practice Exercises vs. Remedial Drill.

II. Criteria for Construction of Remedial Material.

118. Analysis must Precede Diagnosis. 119. Diagnosis must be Specific. 120. Each Identified Skill must be Paralleled by a Valid Drill Unit. 121. The Complete Area of the Skill Must be Covered. 122. The Underlying Skills Must be Synthesized in Remedial Units. 123. Time and Achievement Standards Should Accompany Remedial Drill Material.

III. Summary.

IX. Diagnostic and Remedial Techniques; Arithmetic 231

124. Introduction. 125. Scope of Remedial Teaching in Arithmetic.

I. Identification of Arithmetical Skills.

126. Specific Skills in Arithmetic as Basis for Test and Remedial Units.

II. Remedial Instruction in Arithmetic.

127. Materials for Remedial Work in Arithmetic. 128. Using the Remedial Materials in Classroom.

III. Summary.

X. Diagnostic and Remedial Techniques; Language 251

129. Introduction. 130. A Language Consciousness.

I. Identification of Specific Language Abilities.

131. Analysis of Language Skills. 132. Oral Language Skills. 133. Written Language Skills.

CHAPTER PAGE

II. Diagnosis of Language Disabilities.

134. Diagnosis of Oral Language Disabilities. 135. Diagnosis of Difficulties in Written Language. 136. Diagnostic Exercises in Written Language.

III. Remedial Instruction in Language.

137. Language Diagnosis as a Basis for Remedial Instruction. 138. Remedial Exercises in Language; Spelling. 139. Remedial Exercises in Punctuation. 140. Remedial Exercises in Form and Appearance. 141. Remedial Exercises in Correct Language Usage. 142. Remedial Exercises in Sentence Structure. 143. Remedial Drill in Choice of Words. 144. Drill on Language Organization.

IV. Summary.

XI. Diagnostic and Remedial Techniques; Reading, and other Subjects 285

145. Educational and Social Importance of Reading Ability.

I. Indentification of Major Reading Abilities.

146. Major Objectives in Reading Instruction. 147. Typical Defects in Reading. 148. Oral vs. Silent Reading.

II. Diagnosis and Remedy in Oral Reading.

149. Diagnosis of Oral Reading Difficulties. 150. Remedial Drill for Oral Reading difficulties.

III. Diagnosis and Remedy of Silent Reading Difficulties.

151. Diagnosis of Work-Study Type of Reading; Elementary School Level. 152. Diagnosis of Work-Study Type of Reading; High School and College Level. 153. Remedial Drills for Work-Study Types of Reading Difficulties. 154. Other Remedial Materials in Reading.

IV. Diagnosis and Remedy in Other Subjects.

155. Diagnostic and Remedial Treatment of Other School Subjects.

V. Summary.

CHAPTER PAGE

XII. Tests and the Classroom Teacher 327

I. Summary.

156. Recognition of the Need for Tests in the Classroom. 157. Meaning of Educational Tests. 158. Uses of Educational Tests. 159. Using Tests in the Classroom. 160. Summarizing and Interpreting the Results of Testing. 161. After Testing What? 162. Conclusion.

Appendix 335

A. Distributing Agencies for Tests 335
B. Classified List of Tests and Publishers 337
C. Suggested List of Remedial and Drill Materials 353
D. Illustrative Problems in Educational Measurements. 356
E. Individual Pupil Test Record Form 375
F. Glossary of Terms 376

Index 385

LIST OF TABLES

TABLE NO.		PAGE
1	Distribution of Mental Ability of a 9th Grade Class.	51
2	Marks Assigned by Ten Teachers to an Essay Type Examination in Civics, and a True-False Examination over the Same Material	80
3	Theoretical Basis for Distribution of Objective Test Grades	101
4	Suggested Point Values Corresponding to Class Grades	102
5	Scores on a History Test Taken by an 8th Grade Class	118
6	Arithmetic Scores of Thirty-Seven 7th Grade Pupils.	119
7	Frequency Distribution Based on Data in Table 6	125
8	Relation of Range and Size of Class-Interval	126
9	Distribution of Scores and Calculation of Arithmetic Mean	133
10	Arithmetic Test Scores Arranged in Descending Order	134
11	Arithmetic Scores in Frequency Distribution	136
12	Distribution of Scores on a First Grade Reading Vocabulary Test	138
13	Frequency Distribution	143
14	Standard Deviation from Ungrouped Data	147
15	Standard Deviation from Frequency Distribution	148
16	Scores of a Tenth Grade Class on a Plane Geometry Test	151
17	Pairs of Test Scores	154
18	Illustrating the Computation of XY Products	158
19	Correlation Table	159
20	Relative Ranks	167
21	Table for Assignment of Rank Positions	168
22	Computing Percentiles (Deciles)	170

TABLE NO.		PAGE
23	Interpretation of Percentiles	171
24	Short Method of Computing Percentiles	173
25	Norms and Standards of Accuracy	178
26	Distribution of Scores for Computation of Norms	181
27	Percentile Norms	182
28	Woody-McCall Arithmetic Test Results	186
29	Grade Increments	186
30	Semester Improvement, Monroe Silent Reading Tests	188
31	Distribution of Language Test Scores by Ages	189
32	Age-Equivalents of Composite Scores on Stanford Achievement Examination	190
33	Pupil Record from Compass Diagnostic Test Number 7	192
34	Grade Norms for Compass Diagnostic Test Number 7	192
35	Age-Equivalents of Total Scores; Compass Diagnostic Test Number 7	193
36	Comparison of Units of Measurement	194
37	Grade Levels for Thorndike-McCall Silent Reading Scales	196
38	Thorndike-McCall T-Scores and Reading Ages	199
39	Data for Reading Ages	200
40	Data for Reading Quotients	202
41	Data for Accomplishment Quotients	204
42	Test Record of a Pupil	209
43	Distribution of Practice; Remedial Unit Number 27.	217
44	Analysis of Problem Solving	223
45	Distribution of Practice; Remedial Unit Number 31.	227
46	Scope of Remedial Drill Units in Whole Numbers	243
47	Distribution of Practice in an Addition Test Prepared by a Teacher	248

LIST OF FIGURES

		PAGE
1	Graphic Record Card	57
2	Illustrating the Principle of Sampling	77
3	Mid-points and Limits of Class-Interval	124
4	Illustrating the Principle of Moments of Forces	129
5	Illustrating the Calculation of the Mean	131
6	Illustrating Need for Measures of Dispersion	139
7	Comparison of Sigma and Q	145
8	Graphic Presentation of the Median, 25th, and 75th Percentiles	184
9	Comparing Achievement with Grade Norms	186
10	Illustration of Need for Age Norms	190

LIST OF PROBLEMS

APPENDIX D

I. Tabulation of Test Scores

		PAGE
1	Geography Test Scores of 21 Fifth Grade Pupils	357
2	Reading Test Scores of 41 Fourth B Grade Pupils	358
3	Language Test Scores of 31 Sixth Grade Pupils	358

II. Finding the Arithmetic Mean

4	Computing the Arithmetic Mean from Frequency Distribution	361
5	Computing the Arithmetic Mean from Frequency Distribution	362
6	Computing the Arithmetic Mean	363
7	Computing the Arithmetic Mean	363

III. Finding the Median

8	Finding the Mid-Measure	363
9	Finding the Mid-Measure	363
10	Finding the Mid-Measure	364
11	Tabulating, Finding Median, and Arithmetic Mean	368
12	Tabulating, Finding Median, and Arithmetic Mean	369

IV. Expressing the Variability of Test Scores

13	Finding the Semi-Inter-Quartile Range	370

V. Relative and Absolute Ranks

14	Assignment of Relative Ranks	371
15	Computing Percentile Ranks	371

VI. Using Tables of Norms

16	Finding Age-Equivalents from Test Scores	372
17	Comparing Test Scores with Grade Norms	373
18	Finding Quotients from Ages Given	374

THE USE AND INTERPRETATION OF EDUCATIONAL TESTS

THE USE AND INTERPRETATION OF EDUCATIONAL TESTS

CHAPTER I

INTRODUCTION

The purpose of this chapter is the discussion of the following points which seem to underlie the development of the idea of measurement in education:

a. The idea of measurement in education is not necessarily a recent movement, but was in the minds of a number of pioneers many years prior to its general public acceptance.
b. The real pioneers in the measurement movement.
c. The effect of studies of the inaccuracy of teachers' marks on the measurement idea.
d. The effect of the "school survey" idea on the development of tests in school subjects.
e. The stimulation given to the measurement movement by the organization of research bureaus and kindred agencies.
f. The effect of detailed analysis of skills underlying achievement in certain fields.
g. The development of detailed diagnostic tests based upon the analysis of underlying skills.
h. The minimal principles needed in the use and interpretation of tests.

I. Background of the Growth of the Measurement Idea

1. Educational Measurements not new idea. Standardized tests and the types of information resulting from their use in

the classroom are coming to be almost universally identified with good teaching practice. In many ways teachers have always endeavored to measure the progress of their pupils toward an educational goal and to diagnose revealed defects in instruction. This measurement of progress and the accompanying diagnosis have been in the past largely a matter of observation on the part of the teacher. And yet the recent development of standard tests may be regarded as an extension and improvement of an old practice. However, tests are more precise and exact than the ordinary teacher's examination and thus they accomplish the purposes of measurement much more reliably. They permit the setting up of specific and objective goals of achievement which are based upon the actual attainments of children under typical school conditions. They provide answers to such specific questions as "How much should be expected of a given class or pupil under certain school conditions?" or "How much progress should be made in a given subject in a certain school period?"

2. Real Pioneers in Measurements. The idea lying back of the standardized objective test is relatively simple. It involves mainly the careful selection and repeated use of certain examination exercises by groups of individuals until the typical reaction of these individuals becomes known. This typical reaction later becomes known as norms or standards in terms of which other results can be compared and evaluated. In view of this relatively simple situation it is strange that the idea of measurement has been so long in establishing itself generally in the teaching field. There is evidence to indicate that the idea had been comprehended in part by teachers before it was developed and popularized by modern workers. Professor E. L. Thorndike has called attention [1] to a plan for measuring the achievements of pupils which was used by Rev. George Fisher, headmaster of a school in Greenwich, England, as early as 1864. The instrument used was called a "scale book" in which were assigned values to each degree of pro-

[1] *Journal of Educational Psychology,* Vol. 4, p. 551.

ficiency in the different school subjects. It was naturally somewhat crude in its organization, but it included the germ of many of the ideas which are incorporated in our present-day educational tests. Fisher's work, like that of most pioneers, produced no lasting results because he lived too far in advance of the thought and educational practice of his day. His work appears to be isolated from the development of educational tests in this country.

In this connection, it should be noted that the construction of many of our present-day measuring instruments was not possible until the statistical methods needed in the construction of tests had been carefully worked out. In this phase of test development Sir Francis Galton deserves special mention for his early recognition of the need for more exact statistical tools such as were later developed by Pearson, Spearman, and others. In America the real inventor of the comparative test was Dr. J. M. Rice. Rice, who had been a student in Germany, returned to America with the notion that it should be possible to measure efficiency and compare results in education as in other fields of science. Until after he demonstrated the fundamental truth of his theory, by giving tests in arithmetic and in spelling in a large number of school systems in this country and revealing some startling differences in achievement, very little progress was made in the development of reliable measuring devices. Different tests were used for identical purposes, and very often the relative difficulties of these various tests were never known. How could one city or school system compare its work with another under such circumstances? Rice, by giving the same test to pupils in many different cities, took the first significant step in advance in procuring for teachers more reliable and dependable measuring instruments.

The early development of standardized objective tests for the measurement of pupil achievement was contributed to by workers in the fields of both psychology and school administration. The psychologist made measuring instruments for studying the functions of the human mind. The school

administrator was interested more particularly in questions of courses of study, school organization, methods of instruction, supervision, etc. Rice's contribution to educational measurement was, in the use of tests in certain subjects for answering questions concerning the best course of study, time allotment, methods of instruction, and the like. In doing this he focused attention upon the idea of measuring the achievements of school children.

Stimulated by the work of Dr. Rice, Edward L. Thorndike soon began to experiment with tests and scales, and to preach the need of them. In 1909 he presented his scale for the measurement of Handwriting before the American Association for the Advancement of Science and published it in 1910 in the *Teachers College Record.* Since that time he has produced and has stimulated the production of hundreds of contributions to this field. Discussions and press reviews of the early productions and findings of the pioneers in test making denounced them as foolish and impracticable. However, the last ten or fifteen years have seen this attitude completely changed and objective measurement has become definitely a part of the technique of scientific education.

3. Factors in Growth of Measurement Idea. The production of educational tests has been further stimulated by at least three developments not having to do directly with their construction.

(*1*) The numerous important studies of the accuracy of school marks such as those made by Johnson, Kelly, Miles, Smith, Starch, Meyer, Dearborn, and Gray [2] revealed the fact that school marks are highly subjective,[3] and hence inaccurate. This accumulation of evidence not only proved conclusively that such subjective measures as teachers' marks might be expected to vary widely from teacher to teacher but also that the same teacher's marks would vary from time to time

[2] See *Selected References* at end of chapter.

[3] See *Glossary* beginning on page 376 for meaning of this and other new terms.

and from subject to subject. It was shown that very often the difference between success and failure in a given subject was largely the pupil's fortune in being placed in one teacher's section of the class rather than in that of another. This type of demonstration revealed the need for instruments which would yield more accurate measures of achievement.

(*2*) In the early part of this century a prevailing feeling of dissatisfaction with existing educational conditions coupled with a curiosity and a desire on the part of a few educators to reveal true conditions led to an unusually large number of intensive investigations or surveys of certain of the larger school systems. So much interest in the idea of comparative measurement had been created by Rice that many of the so-called "school surveyors" were turning to tests as suitable instruments for comparison. Thus the survey movement stimulated the construction and use of educational tests, and has itself been largely influenced by the derivation of more objective devices for measuring the abilities of school children. In a number of school surveys, those in charge of the administration of the survey of the school situation felt that the existing tests were unsatisfactory, and devised new instruments to meet their needs. The *Cleveland Survey Arithmetic Tests* represent one of many illustrations of this type of development.

(*3*) The development of educational measurements has been greatly facilitated by the establishment of research organizations. Research bureaus have been organized in many of the larger school systems primarily for the purpose of directing and stimulating the application of educational tests. Similar research bureaus have been established in a large number of state educational institutions, and recognition has been given to the work by the department of public instruction in a number of states. This type of research organization has been very influential in popularizing the use of educational tests. Through them, school administrators, supervisors, and teachers have been made acquainted with tests, and frequently they have been trained in their use. Service

has also been rendered by making them easily accessible, and by standardizing tests devised by other workers. In many instances those in charge of the administration of the various research organizations have been directly responsible for the derivation of many new testing instruments.

4. Rapid Growth of Interest in Tests. There are now tests and measuring devices for practically all subjects in the elementary and high school field, as well as in many branches of higher education. Mental tests and all kinds of prognostic, diagnostic, and placement [4] tests are rapidly increasing in number and improving in quality. No one can predict the ultimate range of application and availability. The advantages of standardized tests and measurements are generally recognized by educators and laymen. On the other hand many of their shortcomings are as thoroughly realized. Tests are means to an end. Their value depends upon certain characteristics innate in the tests themselves, upon the proper utilization of their results for analysis, diagnosis, and the development of an effective follow-up or remedial program.

The recognition of the value of standard tests for diagnostic purposes has done much to popularize their use and to transfer them from the hands of specialists to those of the classroom teacher. When tests were new and still in the experimental stage it was probably well that their administration should be kept in the hands of persons other than the classroom teacher. But now that their value has been demonstrated it is important that teachers master the technique of using tests for the purpose of locating specific weaknesses of individual pupils. It is coming also to be recognized as equally important that teachers shall not only be informed as to methods of diagnosis but that they shall also be able to select and apply the proper preventive and remedial measures.

[4] See *Glossary*, beginning on page 376 of *Appendix*, for further interpretation of meaning of terms.

5. Purpose of this Book. The fundamental purpose of this book is to prepare the teacher and supervisor for effective use of tests as integral parts of enlightened teaching practice. For this purpose the reader must be informed concerning the meaning and general utility of tests, the general rules for carrying out a testing program, as well as for the use of informal tests for examination purposes. Chapters II, III, IV, and V provide a basic treatment of these fields. The general principles for the selection and use of tests are common to most standard tests now in use. Such rules and suggestions as are given here are not intended to take the place of such rules and directions as accompany the various tests, but merely to clarify and supplement them.

The teacher successful in the use of tests in his teaching must be able to secure an accurate interpretation of results. For the purpose of developing these skills the discussion in Chapters VI and VII is presented. The simple statistical procedures such as making a frequency table, the calculation of medians, arithmetic means, quartiles, etc., will be found especially helpful in interpreting results where fairly large classes are concerned. The discussion of derived scores in Chapter VII should also be useful in making an analysis and interpretation of test results. Experience with practical problems of this nature is provided by means of a series of eighteen exercises in the *Appendix*. Those wishing to follow the use of tests in the classroom to its logical conclusion will wish to study with particular care Chapters VIII, IX, X, and XI, in which are presented the possibilities and the practical methods of using test results for diagnosing the special difficulties of pupils, and the inauguration of a preventive and remedial program of instruction.

Those who are more deeply interested and wish to study further into the problems of measurement in education will find the appended bibliography as well as the selected references at the close of each chapter of particular value. The field of educational measurements is so rich and the essential

material is so extensive that it is impossible to compress into the few pages allotted to this book even much of that which is considered by many to be fundamental.

Teachers themselves expert in the technique of learning know that passive reading, while yielding information and appreciation, does not develop easy dependable skill in doing the thing described. To provide the opportunity for the student and teacher actually to master certain of these techniques by doing them, a number of practical problems for drill are given in the *Appendix* of this volume, pages 356 to 374. In addition to these problems a "Work-Book in Educational Measurements"[5] has been prepared as a companion volume for this treatment. In this *Work-Book* the reader may solve actual and practical problems of the type which the classroom teacher and supervisor face. Mastery of this text and a careful working of the forty-five projects in the *Work-Book* will practically guarantee to the reader an actual concrete experience with the major problems of a dynamic testing program calculated to be of the greatest service in the improvement of typical classroom situations.

SELECTED REFERENCES

Ayres, L. P., "History and Present Status of Educational Measurements." 17th Yearbook of the National Society for the Scientific Study of Education, Part 2. Public School Publishing Company, Bloomington, Illinois.

Dearborn, W. F., "Relative Standing of Pupils in High School and in the University." University of Wisconsin Bulletin No. 312.

Grey, C. T., "Variation in the Grades of High School Pupils," Baltimore, 1913. High School Teachers Association of New York City. Articulation of High Schools and Colleges. Reports of the above associations, p. 49.

Johnson, F. W., "A Study of High School Grades." *School Review*, 19: 13–24, 1911.

Kelly, F. J., "Teacher's Marks." Teachers College Contribution to Education, No. 66, p. 11. Columbia University, 1914.

[5] Greene, H. A., "Work-Book in Educational Measurements." Longmans, Green and Co., New York. Form A, 1928, Form B, 1930.

Meyer, Max, "The Grading of Students," Science (N. S.) 28: 243–252.

Miles, W. R., "Comparison of Elementary and High School Grades." University of Iowa Studies in Education, Vol. 1, No. 1, March, 1911.

Smith, F. O., "A Rational Basis for Determining Fitness for College Entrance." University of Iowa Studies in Education, Vol. 1, No. 2. December, 1912.

Starch, Daniel, "Educational Psychology." The Macmillan Company, 1919. Chapter XXII.

Thorndike, E. L., "Handwriting Scales" in *Teachers College Record*, Vol. 2, No. 2. March, 1910.

CHAPTER II

THE MEANING OF EDUCATIONAL TESTS

This chapter presents a discussion of the following points involved in securing a clear notion of the meaning of educational tests:

a. The qualities which distinguish standard tests from informal examinations.
b. The general superiority of standard educational tests over other educational measuring devices.
c. The distinguishing features of tests and scales.
d. The differences between rate tests and power tests.
e. The types of mental examinations.
f. The distinguishing features of general survey and diagnostic tests.

I. What Tests Are

6. Measurement in Education. In general, in any field of human endeavor, the most efficient results are attained when the worker has definite goals toward which to work, and dependable instruments for determining progress. A definite aim enables the worker to direct his efforts toward the particular task to be accomplished. By the proper use of instruments for measuring results it is possible for the worker to know what he has accomplished. Thus a competent test in long division will give to the teacher a measure of his relative success in teaching that subject. Accurate measuring instruments also aid in discovering when emphasis has been misplaced. For example, a child who has been given an unusual emphasis on oral reading may satisfactorily pronounce words

appearing on the printed page, but may be sadly lacking in ability to get meaning from these same words.

A competent silent reading test will reveal this misplacement of emphasis on the part of the teacher. Measuring instruments also make it possible for the worker to learn definitely whether materials and methods are effective by resorting to experimental methods. This is as true in the field of teaching as in other fields. Without specific aims the teacher cannot plan his work effectively. He cannot know, except in an indefinite way, what he is to do. It is equivalent to telling a teacher that he is to start out and walk for a certain period without giving him any notion of which direction he is to take or how far away his destination may be. If, on the other hand, the goals of instruction are clear-cut and accurate, means of determining progress are provided, and the probability of a timely arrival at the goal is greatly increased.

7. Characteristics of Educational Tests. Educational tests differ in several respects from the typical informal examination such as teachers commonly construct. First, the questions or exercises which are used in educational tests are often much more carefully selected to coincide with the purpose for which the test is designed than is true of the ordinary informal examination. For example, a standard test having for its purpose the measurement of ability to locate the states of the United States contains only such exercises as relate specifically to that purpose. The traditional informal examination frequently contains exercises from many fields. Second, the exercises in a standard educational test are commonly arranged in accordance with certain principles of test construction so as to form an accurate measuring instrument. These principles of test construction, being quite technical, are not discussed in this book. A study of these principles may be safely postponed. For the present it is enough for the reader to note that there are principles of arrangement of items within a test which are important. The reader who is particularly interested in such material will find excellent

treatments in Chapters XVII to XX of "Tests and Measurements in High School Instruction," [1] and in Chapter IV of "The Theory of Educational Measurements." [2]

In the third place, the test or scale is standardized, i.e., the test is given to a large number of children of varying age and school classification. From these results the norms or standards are obtained. The interpretation of norms and standards and the meaning of certain scores derived from these values is discussed in Chapter V of this book. A fourth point of great importance, but only recently recognized and applied in connection with test development is the fact that test scores yielded from carefully constructed tests are more readily translatable into specific remedial procedures where needed. This is nothing more than saying that a really valuable test, in addition to giving a cross-section of the situation, must diagnose, and that this diagnosis must be so specific that the results may be readily interpreted in terms of the specific kind and approximate amount of remedial drill needed. A moment's thought will show the reader that a busy teacher can hardly be expected to construct tests possessing the values of standard tests. Neither should the teacher be forced to depend upon his general powers of observation for this information. A clear appreciation of the possibilities of tests as integral parts of superior teaching practice will reveal that the teacher's most important task in this regard is the selection of the tests best suited to his needs, just as it has long been his duty to select the textbooks best suited to his purposes. In a similar way the teacher's relation to preventive and remedial materials and measures should be one of selection and application rather than one of construction.

8. What Tests Do. Probably every alert teacher has at some time earnestly desired to know whether instruction in certain school subjects was particularly superior or inferior, effective

[1] Ruch, G. M., and Stoddard, George D., "Tests and Measurements in High School Instruction," World Book Company, 1927.

[2] Monroe, W. S., "The Theory of Educational Measurements," Houghton Mifflin Company, 1923.

or ineffective. Many teachers have taken steps to answer this highly important question through the use of one kind of measuring instrument or another, yet relatively few classroom teachers are utilizing to the fullest extent one of the most valuable instruments at their command. Many supervisors and administrators are frequently obliged to admit that they did not have adequate data on which to base their decision as to the efficiency of a certain method of instruction but had to be guided by their own personal opinion. With the development of more valid and reliable measuring devices as a part of the scientific technique of education this will become more and more unnecessary.

Standardized tests are not panaceas for all educational inadequacies but unquestionably they are useful in the evaluation of instruction. Standard tests are objective and definite. A test score has meaning and permits of definite comparison and conclusions, while the ordinary class mark or teacher's grade is largely lacking in meaning. Meaning is attached to the test score through the fact that the test itself is standardized. A quality score of 46 assigned to a certain handwriting sample is meaningless until one is informed that 46 points is standard quality for a fourth grade child. If graded as a second grade product such a sample might be assigned an *A*. If graded by an eighth grade teacher it might readily be given a *C*. There is thus no specific goal of achievement which may be definitely expressed in terms of teachers' grade marks. However, standard test scores are objective and do not shift from pupil to pupil and group to group. By their use specific goals of achievement may be set up. Progress may be measured. By means of a battery of tests carefully administered, new students entering the school system for the first time may be properly classified. Pupils needing special types of training may be discovered. Weak spots in the course of study may be pointed out and methods of instruction evaluated.

Experience in the use of tests in many school systems leads to the conclusion that many times there is an intangible

psychological effect resulting from the administration of a series of tests in a school. The experience of the children while taking the test, and the feeling on the part of the teacher that her work is being carefully checked are both valuable motivating forces making for better and more effective teaching and learning situations. However, this is only a by-product of the use of the test and is no reason for allowing the supervisory work to stop at this point.

9. What Tests Do Not Do. It is generally conceded that standard tests are incapable in and of themselves of improving instruction in any subject. Tests merely reveal the situation and do not do anything to improve significantly the condition revealed by the testing. In the long run a test may be considered as an educational motometer. It reveals the temperature of the water in the automobile radiator. It does not change the motor oil, nor add water to the cooling system. If the motor is running hot it merely indicates that fact. This is equally true of tests. The chief service of tests is their power to reveal the strengths and weaknesses of individual pupils or of the class as a whole. This revelation must be followed by the next logical step, the development of a constructive supervisory program. It is not sufficient that weaknesses be merely revealed. They must be corrected by the use of properly constructed drills and remedial exercises.

Several of the different kinds of tests are described and illustrated in the following sections of this chapter.

II. Kinds of Tests

10. Tests and Scales. Educational measuring instruments are usually classified as tests or scales. As a matter of fact this classification is not very sharply defined nor entirely satisfactory since there is confusion arising from the similarity of the terms, *scale* and *scaled*. The term *scale* is used to designate a series of exercises which by means of a somewhat technical statistical procedure have been arranged in a

definite order or position, usually in ascending order of difficulty or value. In a scale each exercise or group of exercises of a specified value or difficulty represents as much greater difficulty or value than the exercise or group just below on the scale as the next following exercise or group is more difficult than the exercise or group just below it on the scale. The term *scaled* means that the exercises in a test have been submitted to a large group of children for their responses and have been arranged in order of difficulty as indicated by these reactions. This is nicely illustrated in the case of the exercises comprising the *Thorndike-McCall Silent Reading Scales,*[3] a few of which are reproduced on page 16. Thus it happens that many tests are scales, but not all scales are tests. Scales often represent sources of test material, as in the case of the *Iowa Spelling Scales,* from which words of known spelling difficulty may be selected to make a test for a special purpose. (See pages 17–18 for a sample from the *Iowa Spelling Scales.*) The term *scale* also is applied to certain devices designed to measure products or qualities. Good examples of scales of this type are the *Ayres Handwriting Scale,* and the *Willing Composition Scale.* In this case the scales are not sources of material for tests, but are in themselves instruments for assisting the teacher in assigning more accurate quality grades to the pupil's production. On this account such scales are often designated as quality scales.

Sample Exercises from Thorndike-McCall Reading Scale for the Understanding of Sentences

These exercises from the *Thorndike-McCall Silent Reading Scales* are reproduced here with permission from the publishers.[4] The exercises in these scales are so arranged that each succeeding exercise becomes uniformly more difficult.

[3] The publishers of all tests mentioned in this book are given on pages 335 to 336 of the *Appendix.*

[4] Bureau of Publications, Teachers College, Columbia University, New York.

Exercises from the beginning and end of the scale (Form 1) are given here to illustrate the arrangement of scaled exercises.

Read this and then write the answers. Read it again if you need to.

On Monday Dick saw a red fox, a gray squirrel and a black snake in the woods. The next day he saw a brown rabbit and five brown mice in the field. He killed the fox and all the mice, but let the others live.

1. What was the name of the boy who saw the mice?.....
.....
2. On what day did he see the mice?.....
3. What color was the fox?.....

Read this and then write the answers. Read it again if you need to.

COLERIDGE

I see thee pine like her in golden story
Who, in her prison, woke and saw, one day,
The gates thrown open—saw the sunbeams play
With only a web 'tween her and summer's glory;
Who, when the web—so frail, so transitory,
It broke before her breath—had fallen away,
Saw other webs and others rise for aye,
Which kept her prisoned till her hair was hoary.
Those songs half-sung that yet were all divine—
That woke Romance, the queen, to reign afresh—
Had been but preludes from that lyre of thine,
Could thy rare spirit's wings have pierced the mesh
Spun by the wizard who compels the flesh,
But lets the poet see how heav'n can shine.

32. What appeared after a similar thing disappeared?.....
.....
33. What word not in the last three lines although in the last six lines, indicates a failure to attain perfect poetic truth?.....
.....
34. State in ordinary English just why the songs lacked perfection.
.....
.....
35. Whose songs were half-sung?.....

The Iowa Spelling Scales

The following sample page from the *Iowa Spelling Scales* is reproduced here with permission from the publishers.[5] This is page 16 of the scale for the eighth grade, and shows the groupings and the difficulties of these words for this grade. For example, the word *phosphorus* in Step 16 will on the average be misspelled by 28% of eighth grade children attempting it. The word *client* in Step 12 will be misspelled by 62% of eighth grade pupils on the average. The complete scale is a very useful source of material for the construction of spelling tests and exercises. There is a separate scale for each of the grades from the second to the eighth.

The Iowa Spelling Scales—Grade VIII

extremely
financial
inquiries
literature
representative

64%

calendar
leisure
opportunity
solicit

63%

annually
practically
semester
superintendent

Step 12—58% [6]

62%

client
convenient
council
immense
permanent
principle

61%

accredited
characteristic
courteous
losing
scientific

60%

analysis
correspondence
mortgage
Sabbath

59%

enthusiastic
lieutenant
unusually

58%

all right
alumni
anticipate
assessment
candidacy
continuous
fundamental
geometry
girlie
physician
possess
thorough

[5] The Public School Publishing Company, Bloomington, Illinois.

[6] The per cents refer to the mid-points of the difficulty values assigned to the steps.

57%
thoroughly

56%
accompanying
acquaintance
auntie
originally
recommendation

55%
anticipating
circuit
disappoint
equipped
immediately

Step 13—50%

54%
confirmation
materially
recommend
thesis

53%
pneumonia
referred

52%
satisfactorily

51%
bulletin
disappointment
grateful
referring

50%
apparatus
co-operating
inconvenience
judgment
rec'd

49%
efficiency
inst.
remembrance

48%
grippe
prior

47%
affectionately

46%
politician
tuberculosis

Step 14—42%

45%
affidavit

44%
ingredients
privilege

43%
guarantee
rheumatism

42%
appendicitis
Hallowe'en

41%
conveniently
enthusiasm
guaranteed
occasionally

39%
accommodate
peculiarities

38%
schedule

Step 15—34%

37%
definitely
fraternally

35%
anniversary

34%
zephyr

33%
chautauqua
X-ray

31%
canvass

Step 16—27%

28%
phosphorus

In many tests the material represents a systematic sampling from a scale or group of scales. It is a common practice for such tests to be comprised of a rather limited number of exercises carefully arranged in ascending order of difficulty. Such tests may be treated as scales by giving the subject credit with the value of the most difficult exercise which he is just able to solve or they may be treated as tests by scoring in terms of the number of exercises solved correctly regardless of difficulty. Thus it must be clear to the reader that the lines of demarcation between tests and scales are not so clear as many seem to believe. Modern practice in test construction tends to move away from the careful scaling of exercises with the corresponding assignment of difficulty values, toward the practice of including large numbers of exercises of unknown value but sampling the field so widely that the factor of difficulty does not decrease the reliability of the test.

Educational tests in general may be catalogued as: (1) tests of intelligence or general mental ability; (2) tests of general achievement; and (3) tests for diagnosing specific difficulties in the various school subjects.

11. Mental Tests. General mental ability is commonly thought of as the power of the individual to learn. It is usually measured in terms of the extent to which he has acquired a number of specific and largely unrelated abilities. Most intelligence tests consist of a battery of different tests measuring many different fields. The total score on the battery is taken as an index of general intelligence. In a sense, general mental ability may be thought of as a cable composed of many strands and fibers of varying size and quality, each representing some particular phase of ability. The test of general intelligence is merely a device for taking a cross section of this cable. To the extent that the particular measuring instrument has in it the elements which reveal the large and important strands of the cable the test is valid and more nearly reveals the true mental status of the individual measured.

12. Individual Mental Examinations. The mental tests which have been developed fall into two general divisions: (1) individual tests; and (2) group tests of mental ability. The individual test can be used to compare the intelligence of different persons, but only when they have had exactly the same learning opportunities, and when they are measured by the same instrument under exactly the same conditions. Tests capable of this type of discrimination are largely patterned upon the original Binet-Simon tests, an individual examination. Binet prepared the first edition of this mental examination in 1905 and revisions appeared rapidly after that. The 1908 and 1911 editions were prepared in collaboration with Dr. Simon. Decroly and Degand published a revision in Belgium in 1910, and the revisions by Goddard in America and Johnston in England appeared in 1911. Kuhlman's revision was published in 1912. The most popular revision of this scale in this country, the Stanford Revision, was published by Terman in 1916. The Herring [7] revision appeared in 1922.

The general procedure in administering the Stanford Revision [8] is representative of that of many of the other revisions mentioned. The type of performance varies considerably with the different exercises. These exercises are presented to the child by means of spoken directions. The examiner tests one child at a time. The test should be given in a quiet room where there is freedom from distraction, and a friendly attitude between examiner and subject should be maintained. The examiner is directed to make sure that the subject understands what is to be done, and in all cases the burden of proof is with the examiner to show that the subject has responded in a way that is representative of his ability. The following brief quotation from Terman [9] emphasizes the importance of

[7] World Book Company.

[8] Terman, L. M., "The Measurement of Intelligence," Houghton Mifflin Company.

[9] Terman, L. M., "The Measurement of Intelligence," Houghton Mifflin Company, 1916; page 124.

the examiner's getting into "rapport" with the subject to be examined. "The examiner's first task is to win the confidence of the child and overcome his timidity. Unless rapport has first been established, the results of the first tests given are likely to be misleading. The time and effort necessary for accomplishing this are variable factors, depending upon the personality of both the examiner and the subject. In a majority of cases from three to five minutes should be sufficient, but in a few cases somewhat more time is necessary."

The chief advantage of the individual mental examination is the opportunity it affords for a trained examiner to observe the child's behavior under standard conditions. Since only one person can be examined at a time the individual mental examination is quite expensive in terms of time. Workers in this field, led by the work of Terman, Otis, Thorndike, Yerkes, and others, and stimulated by the psychological testing during the World War, finally developed very satisfactory instruments for measuring large groups of subjects at the same time. Of course, when group tests are given the difficulties and mistakes of individuals are not so readily observed by the examiner. In the individual examination the examiner may guide the subject into the best possible response for that subject. Along with this advantage there is the disadvantage that individual tests demand the services of highly trained examiners. Training in the field of psychology and mental testing is highly desirable from the standpoint of the administration of the tests, and the interpretation of the results obtained. The significance of training is made clear in the following statement by Terman in his "Measurement of Intelligence" (pages 33–34): "Binet reminds us that the difference between the scientific and unscientific way of doing a thing is not necessarily a difference in the nature of the method; it is often merely a difference in exactness. Science does the thing better, because it does it more accurately."

The point to remember is that the correct interpretation of an individual's reactions under testing conditions is an

expert matter and should be attempted only by a trained person. Not every teacher in a school need become a trained mental examiner. In fact it is probable that only a few are gifted with the patience, tact, and powers of observation necessary to enable them to become successful examiners. Many school administrators who cannot afford a full-time psychological examiner are provided with means by which one or two carefully selected teachers can receive the proper training in this important skill and yield sufficiently expert service to the school system as a whole.

13. Group Tests of Mental Ability. Group tests for measuring intelligence include all those which may be given by one examiner to a number of children at the same time. The outstanding advantage of these measuring instruments in comparison with those which must be administered to children individually is the great economy of time resulting from testing children in groups. There is an added advantage in that less training and experience are required of the examiner. On the other hand, there is the disadvantage of slightly less reliable results from group tests. A few of the more commonly used tests of this type are: the *Terman Group Test of Mental Ability,* the *Otis Group Intelligence Test,* the *Otis Self-Administering Tests,* the *Illinois Examination,* the *Haggerty Intelligence Examinations,* the *Pressey Intermediate Classification Test,* the *Miller Mental Ability Test,* and the *Detroit First Grade Intelligence Test.* Because of the fact that it is representative of a large number of tests of this type Test 1 of the *Detroit First Grade Intelligence Test* is reproduced here with the permission of the publishers. This material represents the examiner's copy of the test. The test itself does not call for reading on the part of the child.

While the foregoing intelligence examinations have been listed together and no attempt has been made to classify them as to function, it must be borne in mind that their functions are not necessarily identical. Some are more appropriate for

Detroit First Grade Intelligence Test [10]

TEST 1. Look at the top row of pictures. Mark the thing that grows on a tree.

*

Look at the next row. Mark two things we cut with.

* *

Look at the next row. Mark three things made of wood.

* * *

Look at the next row. Mark four things that lay eggs.

* * * *

use in the high school than others. On the other hand, certain ones are specifically intended for use in the elementary school.[11]

14. Special Aptitude and Prognostic Tests. If the general concept of the measurement of mental ability is accepted, then

[10] From Examiner's Manual for Detroit First Grade Intelligence Test, Copyright, 1921, 1926, by World Book Company, Publishers, Yonkers-on-Hudson, New York.

[11] More exact information concerning these tests and others named in this and succeeding chapters may be obtained by studying the catalogues of the chief publishers of tests. A list of test publishers and distributors is given on pages 335–336 of the *Appendix*.

it is possible to develop instruments for the measurement of specific types of intelligence. Devices such as the *Stenquist Mechanical Aptitude Tests, O'Rourke Mechanical Aptitude Tests,* the *Rogers Tests of Mathematical Abilities,* the *Seashore Musical Talent Tests,* and the *Iowa Placement Examinations, Aptitude Series* in each of the fields of Chemistry, English, Foreign Language, Mathematics, and Physics are supposed to measure the individual's capabilities to learn in some definite field. Such instruments may be called specific intelligence tests. Tests of this kind have also been called prognostic tests.

The accompanying reproduction of a few of the exercises comprising three of the four parts of the *Mathematics Aptitude Test* (MA-1) of the *Iowa Placement Examinations* may serve to give a more definite idea of the content of such tests. It will be noted that aptitude tests are designed to measure specific underlying abilities and not specific achievement in the selected field. In the complete test of which this is a portion, Part 1 has 15 exercises, Part 2 has 15 exercises, Part 3 has 20 exercises, Part 4 (Reading of mathematical material) has 15 exercises, making the total possible score 65 points. The test requires a total working time of 40 minutes.

Specimen Exercises from Mathematics Aptitude (MA-1) Test of Iowa Placement Examinations [12]

PART 1

Directions: Each of the following number series is made up according to a rule. Discover the rule for each example and write the next *two* terms on the dotted lines. You have 5 minutes for Part 1.

Sample: x, $2x$, $4x$, $8x$	.....$16x$.....	.....$32x$.....
1. 64, 32, 16, 8,	.....	.....
2. 8, 0, 7, 0, 6, 0,	.....	.....
3. $\frac{9}{25}$, $\frac{13}{21}$, $\frac{17}{17}$,	.....	.....
4. $\frac{2\cdot4}{1\cdot3\cdot5}$, $\frac{2\cdot4\cdot8}{1\cdot3\cdot5\cdot7}$,	.....	.....

[12] Copyright, 1925, University of Iowa. Published by Bureau of Educational Research and Service, Iowa City.

PART 2

Directions: Place the answer to each question on the dotted line at its right. Use the margins of this page for figuring. You have 10 minutes for Part 2.

Sample: A circle is revolved about a diameter as an axis. What geometrical figure is formed? ....a sphere....

	Answers
1. If a rectangle is revolved using a side as an axis what figure is generated?	.....
2. A box contains 10 black balls and 20 white balls. A man draws out 9 balls at random. How many are probably white?	.....
3. A certain polygon has $n+1$ vertices. How many sides has it?	.....
4. If x is the sum of the base and the altitude of a rectangle, what dimensions of the rectangle will give a maximum area?	.....

PART 3

Directions: Below are 20 examples, each consisting of two statements. *You are to assume that the statement beginning with the word " given " is true.* Read the first statement in each example, and then examine the " conclusion." Decide whether the conclusion is true or false. If the conclusion is true, place a T after it on the dotted line; if the conclusion is false, place an F after it on the dotted line. *Do not guess.* You have 10 minutes for Part 3.

Sample:

Given: A square and a triangle have the same base and altitude.
Conclusion: Therefore the square is larger than the triangle. ..T..

1. *Given:* A is greater than B, B is greater than C. *Conclusion:* Therefore A is greater than C.	1.....
2. *Given:* M^2 is less than N. *Conclusion:* Therefore M is less than N.	2.....
3. *Given:* A times B equals C times D. *Conclusion:* Therefore A plus B equals C plus D.	3.....
4. *Given:* a is greater than b. *Conclusion:* Therefore the mean of a and b is less than a.	4.....

15. Tests of Achievement. Tests of achievement, or measures of what the pupil has learned as a result of his school or other type of educational experience are usually classified according

to their form or function. In the matter of form most achievement tests are classified as rate or power tests, and quality or product scales. According to function they are classified as survey or diagnostic tests.

Rate Tests. Rate tests are measures of the speed with which a pupil is able to respond to certain standardized activities of a uniform nature. Usually the exercises are relatively easy and there is little or no question about the pupil's ability to do them. Rate tests also usually consist of exercises approximately equal in difficulty, and the time allowed is such that ordinarily no pupil is able to finish the test. The number of exercises finished within the time limit is taken as the pupil's rate of work. The quality of his performance may be expressed by the per cent of exercises done correctly. Such a test implies that our educational objectives are to train children to extend their power over a larger and larger area of content of constant difficulty. The following are among the most widely used rate tests: *Monroe Standardized Silent Reading Tests Revised;* the *Cleveland Survey Arithmetic Tests; Courtis Standard Research Tests in Arithmetic, Series B; Burgess Picture Supplement Scale for Measuring Reading Ability;* the *Courtis Silent Reading Test No. 2;* and the *Chapman-Cook Speed of Reading Tests.*

The fundamental features of rate tests in general may be seen in the reproduction of a portion of the *Chapman-Cook Speed of Reading Tests* which is given below.

Specimen Exercises from Chapman-Cook Speed of Reading Test [13]

SPEED OF READING TEST

PRELIMINARY DRILL

TEACHERS' DIRECTIONS—The material given below is used to give the pupils some preliminary exercise so that they will understand precisely what has to be done. All tests of speed demand exact and immediate understanding of the directions. You cannot be too

[13] The samples of the *Chapman-Cook Speed of Reading Test* are reproduced with the permission of the publishers, The J. P. Lippincott Company, Philadelphia, Pa.

careful to see that every pupil gets the idea of the test before the papers are turned and the timing commences.

In the second half of each of the following sentences, **one word** spoils the meaning of the paragraph. Find this word as quickly as you can, and cross it out. You must not write anything. Cross out the word that should not be there. In the first paragraph you can see that the word "thinnest" does not fit in with the meaning of the rest of the paragraph. You will therefore take your pencil and draw a line through the word. In the second paragraph, "nail" is wrong; you must, therefore, cross it out.

It was such a cold, boisterous, and wintry day that every person who was walking wore the thinnest clothes that he could find in his clothes-closet at the time.

There was a fire last night, and five houses were burned to the ground. It all happened because someone was careless, and threw a nail into the waste-paper basket.

Now cross out the one word which should not be there in the six paragraphs below. After you have done them, the teacher will go over them orally to be sure that you understand.

A. There was not a drop of ink in the house, for someone had broken the bottle we kept it in, so Mary decided to finish her letter with a pen.

B. Yesterday I went down town to buy some shoes and rubbers, but, when I got home, I found I had forgotten to go to the flower-store to get them.

C. The water had frozen, making the road as slippery as glass. It was only with the greatest difficulty that I prevented myself from fighting as I made my way home.

D. I was not in time for school, because I played marbles on the way; so the teacher sent a note to my parents, saying I had been early that morning.

E. One of the boys was extremely rude to the teacher so she made him come and stand by her desk, to show that he had been a very good boy.

F. When I am enjoying anything very much, time seems to go very quickly. I noticed this the other day, when I spent the whole afternoon reading a very uninteresting book.

TEACHERS' DIRECTIONS—Do not proceed to the test proper until you are quite certain that every pupil understands exactly what has to be done. Answer any questions that may arise, and just before commencing make one of the pupils state definitely just what is required.

This test has been constructed to make marking as easy as possible. After the test has been taken the pupils may change papers and do the marking while the teacher dictates the correct answers. Score is the number right.

On the other side of this sheet, there are thirty paragraphs like the ones you have done. When you are given the signal, and not before, turn the sheet over, and cross out in the **second** part of each paragraph, the **one** word which should not be there. Ask no questions. Do not stay too long on any one paragraph, but go on to the next. Remember you do no writing; you merely cross out the **one word** in the second part of the paragraph.

This is a test for speed and accuracy. Be sure to work as fast as you can, and yet not make mistakes.

Allow exactly two and one half minutes (2½ mins.)

Cross out, in the second part of each paragraph, the one word which spoils the meaning. (2½ minutes allowed)

1. Tom got badly hurt the other day, when fighting with his older brother. As soon as this happened, he ran home to his mother, laughing as hard as he could.

2. The woman burned herself badly, while she was making soup for our Sunday dinner, just because she was so careless as to tip over a pan of cold water.

3. One day we found something in the woods that had four legs, long sharp teeth, and thick brown fur, so we knew it must be a bird of some sort.

4. We started to cut down a tree in our front yard, but, after working for two hours and a half, we gave it up, because our hammers were no good.

16. The other day we forgot all about our little kitten, and were gone all day. When we returned, we found him on the back porch, barking to be let in.

17. We are interested in what is happening in America and the other great nations of the world, so each morning we read the pages of the dictionary with great care.

18. The river was so wide and the current so swift, that we always kept a boat, with oars and a sail ready for use, whenever we wanted to walk across.

19. Mary said she mailed a letter to me at the post office the day before yesterday, but the milkman did not leave anything for me when he came this morning.

Power Tests. Power tests or the so-called scaled tests (see page 15 of this book for a discussion of this point) measure a pupil's ability to do more and more difficult exercises within a given field of subject matter. In such tests achievement is expressed in terms of the difficulty of the exercise or activity which the pupil is just able to perform. If *rate tests* may be compared to running a race in which a series of hurdles of uniform height are to be cleared, *power tests* may be compared to a race in which the hurdles are very low at the start but become gradually higher and higher as the race goes on until no one is able to clear them. In the first case, the achievement (rate score) would be expressed in terms of the number of hurdles cleared in a specified time. In the second case, the score (ability) would be expressed as the height of the highest hurdle which the individual was just able to clear.

A power test consists in a series of exercises within a given field of subject matter arranged in ascending order of difficulty, and is intended to measure a pupil's power to do more and more difficult tasks of the type included in the test. Usually no measure of the pupil's rate of work is secured because the time allowed in such a test is more than enough for all subjects to complete as many of the exercises as they are able to do. Sometimes, however, rate and quality are combined by taking as the pupil's score the number of exercises of increasing difficulty (or value) done correctly in the specified time. Theoretically a pupil's score on a power test is the degree of difficulty of the most difficult exercises he is able to do with a specified degree of accuracy. Such a score is laborious to compute from a pupil's performance, and for this reason the number of exercises done correctly is generally taken as his score. A few of the most common measuring instruments of this type are: *Van Wagenen Reading Scales, Woody Arithmetic Exercises, Thorndike-McCall Reading Scales, Hotz Algebra Scales,* and the *Henmon Latin Tests.*[14]

The characteristics of power tests are illustrated in the

[14] A list of publishers of these tests is given in the *Appendix.*

reproduction of the *Woody-McCall Mixed Fundamentals Tests* in Arithmetic which appears herewith:

Specimen of Survey Test in Arithmetic [15]

WOODY-McCALL MIXED FUNDAMENTALS: FORM I

Name..... *Age*..... *Grade*..... *Building*..... *City*.....

Get the right answer to as many examples as you can in 20 minutes. Do all work on the front or back of this sheet.

(1) Add: 2 / 3

(2) $2 \times 3 =$

(3) $3\overline{)6}$

(4) Subtract: 2 / 1

(5) Multiply: 23 / 3

(6) Subtract: 13 / 8

(7) Add: 17 / 2

(8) $3 + 1 =$

(9) Subtract: 16 / 9

(10) Multiply: 254 / 6

(11) $4 \div 2 =$

(12) Add: 23 / 25 / 16

(13) Subtract: 393 / 178

(14) $2\overline{)13}$

(15) Add: 9 / 24 / 12 / 15 / 19

(16) Multiply: 5096 / 6

(17) $2\frac{3}{4} - 1 =$

(18) Add: \$12.50 / 16.75 / 15.75

(19) Multiply: 7898 / 9

(20) $\frac{1}{4}$ of 128 =

(21) Add: 547 / 197 / 685 / 678 / 456 / 393 / 525 / 240 / 152

(22) Multiply: 287 / .05

(23) $248 \div 7 =$

(24) Subtract: 27 / $12\frac{5}{8}$

(25) Add: 4.0125 / 1.5907 / 4.1000 / 8.6730

(26) Multiply: 9742 / 59

(27) $\frac{7}{8}$ of 624 =

(28) Add: .49 / .28 / .63 / .95 / 1.69 / .22 / .33 / .36 / 1.01 / .56 / .88 / .75 / .56 / 1.10 / .18 / .56

(29) $\frac{1}{8} \times 2 =$

(30) Multiply: $987\frac{3}{4}$ / 25

(31) $\frac{3}{4} \div 5 =$

(32) $7.3 - 3.00081 =$

(33) $9\overline{)69 \text{ lb. } 9 \text{ oz.}}$

(34) Multiply: $.0963\frac{1}{8}$ / .084

(35) $25.091 + 100.4 + 25 + 98.28 + 19.3614 =$

[15] Reproduced with the permission of the publishers, Bureau of Publications, Teachers College, Columbia University, New York City. Copyright, 1920 by Teachers College.

16. Quality and Product Scales. *Quality Scales.* Achievement may also be expressed in terms of quality scales. In such subject matter fields as handwriting, composition, etc., the measuring instruments must obviously be of a different type. In these subjects the pupil produces a performance which cannot be considered as either right or wrong or at least as partially right. The problem becomes one of describing its degree of quality. For this purpose "quality scales" have been constructed which consist of a series of specimen performances of the type to be described. These exhibit varying degrees of quality from the lowest quality to the highest. The specimens are arranged systematically in order of increasing quality. Usually the quality of each is described numerically. A quality scale of this type is used by matching the performance to be described with the specimen of the scale which most nearly resembles it in quality. The following quality scales have been widely used: *Willing Scale for Measuring Written Composition; Ayres Handwriting Scale, Gettysburg Edition; Hillegas Scale for Measuring Written Composition;* the *Thorndike Handwriting Scale;* and *Knight-McClure Arithmetic Neatness Scales.*

Product Scales. In product scales the exercises or materials are equally spaced on a scale of difficulty, and all exercises of a given difficulty or value are grouped together. On this account product scales represent excellent sources for rate or power tests. The *Ayres Spelling Scales,* the *Iowa Spelling Scales,* and the *Simmons High School Spelling Scales,* and the *Hahn-Lackey Geography Scales* are examples of this type of material.

A point of more than passing interest and significance to the teacher is the recent tendency of authors of modern textbooks to include in their books large amounts of such source material for drill and testing purposes.

17. Survey vs. Diagnostic Tests. *Survey Tests.* The majority of tests which have been devised for school use have for their function the measurement of achievement, or what the

pupil has learned. They are accordingly called achievement tests. Many achievements appear to be specific. In such fields as the operations of arithmetic a large number of skills which are relatively independent of one another are recognized. For example, a pupil may attain a very high degree of achievement in the addition of integers without attaining anything like an equivalent degree of achievement in the addition of fractions. In fact, each of these achievements, in turn, is general, in a sense, for it has been shown that a pupil may attain one degree of achievement in doing long-column additions, and exhibit a distinctly different degree of achievement in short-column addition.

This suggests the possibility of at least two types of achievement tests; first, those which measure certain general achievement; and second, those which yield independent measures of a number of specific achievements within a given field. As illustrations of this first type almost any one of the several hundred achievement tests may be mentioned. Such tests as the *Compass Survey Tests in Arithmetic,* the *Courtis Series B Arithmetic Tests,* the *Woody-McCall Mixed Fundamentals in Arithmetic,* the *Thorndike-McCall Silent Reading Scales,* the *Wilson Language Error Test* have had extensive use in the elementary school field. Among the more popular batteries of achievement tests for the elementary school use are the *Stanford Achievement Tests.*[16] Other achievement tests such as the *Briggs Form Test in Language, Kirby Grammar, Van Wagenen's Reading Scales in Literature and in History, Hotz Algebra Scales, Minnick Geometry Tests, Schorling-Sanford Geometry Tests, Iowa Physics Tests, Powers Chemistry Tests, Ruch-Cossmann Biology Tests,* and the *Ruch-Popenoe General Science Tests* have also had extensive use in the high school field.[17]

The characteristics of survey tests designed for use in a single subject matter field are shown in the accompanying

[16] Kelly, Ruch, and Terman, "Stanford Achievement Tests," World Book Company.

[17] See pages 335 to 336 for list of publishers of these tests.

arithmetic examples reproduced from the *Compass Survey Tests in Arithmetic, Elementary Examination,*[18] for Grades 2, 3, and 4. These tests are in four parts, each part consisting of a total of 25 to 35 exercises of the types shown herewith.

Exercises from Compass Survey Tests in Arithmetic, Elementary Examination

PART 1—ADDITION

(Time allowance: 6 minutes)

Add:

								7+3=		1+4+4=		
1	0	4	7	6	3	8	5					
6	3	6	0	3	8	7	9	6+6=		6+0+5=		
—	—	—	—	—	—	—	—					
												8
								8	8	9	7	7
6	9	6						7	7	5	8	9
9	2	3	5	7	3	6	7	5	3	2	1	6
0	7	3	13	19	29	38	45	1	4	5	3	0
—	—	—	—	—	—	—	—	—	—	—	—	—
						95						97
						56			993			972
						88			458			52
70			44			12			120			1010
58			28			73			671			530
—			—			—			—			—

[18] These exercises are from the *Compass Survey Tests in Arithmetic,* copyrighted and published by Scott, Foresman and Company, Chicago, and are reproduced with the publisher's permission. Part 1 consists of 30 examples; Part 2 of 25 examples; Part 3 of 35 examples; and Part 4 of 25 examples.

PART 2—SUBTRACTION

(Time allowance: 5 minutes)

Subtract:

7	6	4	8	11	13	12−5=	37	63
2	1	0	1	4	9		24	50
—	—	—	—	—	—	11−7=	—	—

83	49	77	64	107	990	653	623	400
63	17	30	59	98	628	364	225	181
—	—	—	—	—	—	—	—	—

1020	1330	1025	7053	1315	18906
906	1184	986	6048	739	9927
—	—	—	—	—	—

PART 3—MULTIPLICATION

(Time allowance: 8 minutes)

Multiply:

6	4	7	4	4	8	4	9	1×8=
1	8	1	0	9	0	1	6	
—	—	—	—	—	—	—	—	7×9=

2	9	5	6	9	6	8	20	9×8+6=
9	8	1	9	9	3	2	0	
—	—	—	—	—	—	—	—	8×7+4=
								8×8+8=

10	81	410	204	340
4	3	2	3	5
—	—	—	—	—

301	92	57	16	62
97	15	76	50	28
—	—	—	—	—

310	875	3975	8264
186	589	3042	467
—	—	—	—

PART 4—DIVISION

(Time allowance: 6 minutes)

Divide:

$2 \div 2 =$	$5 \div 1 =$	$6\overline{)37}$	$1\overline{)0}$	$7\overline{)7}$	$2\overline{)18}$
$8\overline{)75}$	$6\overline{)126}$	$3\overline{)141}$	$5\overline{)150}$	$4\overline{)347}$	$9\overline{)630}$
$6\overline{)354}$	$3\overline{)627}$	$2\overline{)1146}$	$5\overline{)3445}$	$4\overline{)1648}$	$9\overline{)8408}$
$8\overline{)24320}$		$7\overline{)19502}$	$8\overline{)25640}$		$71\overline{)2414}$
$42\overline{)126252}$			$125\overline{)4750}$		$29\overline{)13369}$

A few exercises from a number of tests comprising the *Stanford Achievement Test, Advanced Examination,* are reproduced here for the purpose of giving an illustration of one of the better and more comprehensive achievement test batteries.

Specimen Exercises from the Stanford Achievement Tests [19]

Form A

TEST 1. READING: PARAGRAPH MEANING

(29 Exercises)

Sample: Dick and Tom were playing ball in the field. Dick was throwing the ball and ..... was trying to catch it.

Write JUST ONE WORD on each dotted line.

1. Fanny has a little red hen. Every day the hen goes to her nest and lays an egg for Fanny to eat. Then she makes a funny noise to tell Fanny to come and get the .....
2. A kitten can climb a tree, but a dog cannot. This is very lucky for Nellie's kitten. Every time Joe's big dog comes along the kitten climbs a tree and the ..... cannot follow.
3. Anna had never seen a squirrel in her life, although she had always wanted to very much. One day when she was playing under a tree she heard a funny little noise over her head. She looked up, and what do you think she saw? Up there in the ..... was the very thing she had always wanted to see, a .....

TEST 2. READING: SENTENCE MEANING

(80 Exercises)

Samples: Can dogs bark?.....Yes No

Does a cat have six legs?.....Yes No

Read each question and draw a line under the right answer.

1. Is milk white?.....	Yes	No	1
2. Do we sleep in beds?.....	Yes	No	2
3. Is the day as dark as night?.....	Yes	No	3
4. Is green a color?.....	Yes	No	4
5. Is smoke always yellow?.....	Yes	No	5
6. Do men and women dress just alike?.....	Yes	No	6
7. Do ships sail on the sea?.....	Yes	No	7
8. Are all chimneys made of brass?.....	Yes	No	8
9. Are rocks hard?.....	Yes	No	9
10. Is everybody as huge as a giant?.....	Yes	No	10

[19] Copyright 1922, 1925 by World Book Company, Publishers, Yonkers-on-Hudson, New York.

TEST 3. READING: WORD MEANING

(85 Exercises)

Samples: Bread is something to catch drink eat throw wear

A robin is a bird cat dog girl horse

In each sentence draw a line under the word that makes the sentence true.

1. March is the name of a **day food month week year** 1
2. A fat person is always **bad blue cold heavy little** 2
3. A thing that is perfect is always **close early hard little right** .. 3
4. A farmer often raises **bears corn gold paper pictures** 4
5. Cotton is **cool dark heavy soft sweet** 5

TEST 4. ARITHMETIC: COMPUTATION

(47 Exercises)

Get the answers to these examples as quickly as you can without making mistakes.

Look carefully at each example to see what you are to do.

Begin here.

(1)	(2)	(3)	(4)	(5)
		Add	Add	Add
3+2=	3+4=	2	7	13
		5	4	2
		—	—	—

TEST 5. ARITHMETIC: REASONING

(40 Exercises)

Find all the answers as quickly as you can.

Write the answers on the dotted lines.

Use the blank sheets of paper to figure on.

Begin here.

1. How many are 3 eggs and 2 eggs? *Answer* .....
2. Mary is 7 years old. How old will she be in 3 years? *Answer* .....
3. A hen had 9 chicks and 3 of them died. How many were left? *Answer* .....
4. Milk costs 8 cents a pint and the milkman is going to raise the price 2 cents. What will it then cost? *Answer* .....
5. If you buy a pencil for 4 cents and pay for it with a dime, how much change should you get? *Answer* .....

TEST 6. NATURE STUDY AND SCIENCE

(95 Exercises)

Samples: The number of cents in a dollar is **200** **100** **300**

Our rain comes from the **clouds** **moon** **stars**

Draw a line under the word that makes the sentence true.

Begin here.

1. Thanksgiving comes in **July** **January** **November** 1
2. The earth is shaped most like a **baseball** **football** **pear** 2
3. A sweet-smelling flower is the **daisy** **poppy** **rose** 3
4. The month before July is **May** **June** **August** 4
5. The axle is a part of an **ax** **typewriter** **wagon** 5

TEST 7. HISTORY AND LITERATURE

(95 Exercises)

Draw a line under the word that makes the sentence true.

---

1. An elf is a kind of **animal** **brownie** **dragon** 1
2. "The Glass Slipper" reminds us of **Ali Baba** **Cinderella** **Goldilocks** 2
3. The first President of the United States was **Adams** **Jefferson** **Washington** 3
4. The shepherd boy who became king was **David** **Saul** **Solomon** .. 4
5. Columbus made his first voyage to America in **1492** **1620** **1776** . 5

TEST 8. LANGUAGE USAGE

(60 Exercises)

Samples
Apples $\frac{\text{is}}{\text{are}}$ good.
He $\frac{\text{told}}{\text{telled}}$ me.

1. I $\frac{\text{calculate}}{\text{expect}}$ to go soon.
2. Last year uncle $\frac{\text{gave}}{\text{give}}$ me a pair of skates.
3. His leg was $\frac{\text{broke.}}{\text{broken.}}$
4. They have $\frac{\text{gone}}{\text{went}}$ to town.
5. He isn't $\frac{\text{any}}{\text{no}}$ better than you.

Diagnostic Tests. A test which yields separate measures of specific abilities is called a diagnostic test, because it points out or diagnoses the specific weakness of pupils. Diagnostic tests are designed to identify particular strength or weakness on the part of an individual child or of a class as a whole. The location of defects in individuals and in classes is one of the most important and helpful teaching procedures. By far the greater number of standard tests that have appeared are of the "survey type," i.e., they are designed for taking "cross sections" of general abilities rather than for the isolation of specific difficulties and skills. Tests possessing genuine diagnostic features are of most value to the teacher.

A test to be genuinely diagnostic must be valid; i.e., it must cover all of the separate skills which are most important and which according to the aims of the course are to receive the most emphasis. The test must be reliable; i.e., each of the separate skills must be measured accurately enough to stand alone. Each skill must be sampled widely enough so that the results furnish an accurate picture of the ability of the child, his strengths and weaknesses.

A test to be of real instructional value must not only diagnose, but must also point the way to specific remedial and corrective material based on the results of the testing. This has been the weakness of educational testing of the past. Recently efforts have been made to remedy this deficiency by the development of diagnostic instruments with real analytical power, and by the preparation of remedial practice material paralleling the diagnosis. In this field lies the future of testing, and in it lies the possibility of a complete revolution in testing procedure. Real diagnostic instruments and real remedial material will place the emphasis upon analysis immediately following the completion of the teaching of a specific unit of instruction rather than follow the lines of ordinary testing programs. Instead of testing two or three times each school year the teacher will test immediately upon the completion of a subject unit. Diagnosis will reveal the specific weaknesses of the individual pupil and the use of remedial

material will correct the defects without delay. In this way the pupil may be made aware of his own weaknesses and held largely responsible for overcoming them.

The term "diagnostic" as applied to tests has resulted in many misconceptions. Fundamentally all tests may be considered diagnostic in that they yield specific information about pupil achievement. However, the diagnosis afforded by many tests is at least extremely general. Some of the tests in common use that are either diagnostic in their character or are labelled as such by their authors are the *Charters Diagnostic Language Test, Monroe Diagnostic Tests in Arithmetic, Spencer Diagnostic Arithmetic Tests, Los Angeles Diagnostic Tests in Arithmetic, Barr Diagnostic Test in American History, Pressey-Conkling Diagnostic Tests in English Composition, Wisconsin Inventory Tests in Arithmetic, Freeman's Handwriting Scale, Kwalwasser-Ruch Test of Musical Accomplishment, Douglass Diagnostic Tests for Elementary Algebra.*[20] Probably the most extensive diagnostic tests developed in recent years are the *Compass Diagnostic Tests in Arithmetic.*[21] (See pages 42–43 for a sample of the content of Test 1 (Addition) of this series.) This is a series of 20 tests covering over 90 of the most important arithmetic functions. In addition to their diagnostic powers, remedial practice exercises paralleling the diagnosis are available making it possible for the teacher to proceed at once to the correction of the specific defects revealed by the diagnosis.

Certain of the general function tests in use undoubtedly possess some diagnostic power. For example, the *Stanford Arithmetic Tests* of the *Stanford Achievement Tests* are diagnostic for computation and general problem solving within limitations. The *Stevenson Arithmetic Reading Tests* are diagnostic for certain functions of problem solving. The *Haggerty Silent Reading Examinations* and the *Stanford*

[20] See pages 335–336 for list of publishers of these tests.

[21] Devised by Ruch, G. M., Knight, F. B., Greene, H. A., and Studebaker, J. W. Published by Scott, Foresman and Company. (Exercises reproduced with the permission of publishers.)

Reading Tests are also diagnostic within the limits of vocabulary, sentence meaning, and paragraph meaning. However, the main usefulness of these tests lies in the field for which they were designed, that of general survey use.

The distinctive differences between the general survey test and the genuine diagnostic test can be readily shown by an illustration taken from the field of arithmetic. The accompanying problem in long division is typical of items found in many general tests in this field. For a correct response to this problem the child is ordinarily given one point. For an error he receives no credit, and no opportunity is provided for the location of the point at which his mastery of the situation breaks down. On the other hand, suppose that the child, in addition to working this example were to be confronted also by a series of exercises comprising the component parts of complete mastery of long division such as the basic division combinations, simple short division, short division with carrying, the multiplication, addition and subtraction involved in division, and the estimation of quotient figures. The accompanying tabulation which is based upon an analysis of the above long division problem gives a definite idea of the extent to which this detailed analysis of skills may be carried with profit to the teacher as well as the pupil. It will be seen that instead of being worth only one point when correctly solved such a problem is worth twenty-one points since there are at least that many places at which mastery of the skill may break down. Furthermore, in the case of error, this type of detailed analysis permits the exact location of the spot in the work at which the error took place. (See page 44 for a list of the skills involved in the above long division example.)

$65\overline{)22555}$

Specimen of Diagnostic Test in Arithmetic [22]

Standard Mathematical Service

COMPASS DIAGNOSTIC TESTS IN ARITHMETIC

RUCH—KNIGHT—GREENE—STUDEBAKER

EDITED BY G. W. MYERS

TEST I: ADDITION OF WHOLE NUMBERS: FORM A

Name_____ Grade_____ Boy or girl?_____

Age_____ When is your next birthday?_____ How old will you be then?_____

School_____ Date_____

(Name) (City) (State)

Summary of Pupil's Score	Part 1	Part 2	Part 3	Part 4	Part 5	Total
Scores on Parts of Test						
Educational Age Equivalent						
Grade Equivalent of Score						

PART 1—BASIC ADDITION FACTS

Add:

0+2=	4+1=	0	9	5	2	3	1	7	9	8	5	7	6	4	2
1+7=	5+3=	2	1	0	3	4	1	5	2	4	4	4	2	1	6
2+2=	1+0=														
7+6=	6+4=														
2+5=	3+4=	9	4	8	0	4	2	6	5	2	4	3	7	8	8
2+1=	2+7=	5	0	6	7	2	5	6	3	4	5	7	3	3	8
0+9=	1+6=														
9+2=	3+3=														
3+9=	7+4=	9	0	1	9	2	7	1	7	1	7	5	8	3	6
5+8=	0+0=	6	9	7	7	8	6	8	8	6	7	9	0	9	3
9+1=	9+9=														
8+5=	7+5=														
2+8=	9+4=	8+3=		7+2=											

*Score on Part 1 = Number right =*_____
[Total possible score = 70 points]

PART 2—HIGHER DECADE ADDITION

Add:

2	1	6	8	18	4	9	7	24	10	6	16	14	0	4	24	21	4
20	21	22	13	6	12	29	25	7	2	12	6	7	10	42	2	4	10
32	1	4	38	4	33	7	1	11	5	15	0	1	33	40	2	1	5
6	18	16	7	20	3	13	24	9	10	9	12	11	5	2	25	14	28
31	6	9	2	24	19	8	9	32	10	3	7	0	2	6	7	19	9
9	11	20	36	9	4	16	27	2	7	18	26	31	26	13	12	2	23

[22] Copyright 1925, Scott, Foresman and Company.

PART 2—Continued

Add:

19+7=	14+2=	10+3=	20+7=	28+7=	0+20=
12+2=	11+7=	18+8=	26+6=	18+7=	22+ 0=

Score on Part 2 = Number right = ...
[Total possible score = 66 points]

PART 3—COLUMN ADDITION

Add:

4	7	6	9	6	9	9	4	7	8	2	6
7	4	6	4	4	7	2	2	6	0	4	3
5	9	9	6	0	8	4	9	7	1	3	5
	1	3	4	7	9	7	7	4	2	8	7
		4	9	5	1	5	8	8	4	1	0
				0	6	9	8	7	1	8	4
							8	6	2	6	7

Score on Part 3 = Number right × 5 = ...
[Total possible score = 60 points]

PART 4—CARRYING IN COLUMN ADDITION

Add:

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
29	98	17	47	117	766	27	9132	7027	162	653		
47	13	13	4	960	982	96	5112	249	4914	5156		
		67	13	517	932	77	2343	7191	9737	7360		
			49		448	84	7417	9235	67	2661		
			29			98		9059	9280	4099		
						17			427	9730		
						26				9222		

12. Copy and add: 62+604+827+797+997=? (Put your work under 12 above.)
13. Copy and add: 76, 64, 60, 97, 35, and 20. (Put your work under 13 above.)

Score on Part 4 = Number right × 10 = ...
[Total possible score = 130 points]

PART 5—CHECKING ANSWERS IN ADDITION

Directions: Some of the printed answers below are right, and some are wrong. Check each sum by adding downwards. Write your check answer on the line at the top of the example. The first one is already done correctly.

797							
213	165	6566	9270	7774	887	162	653
128	923	3157	9984	9447	756	4914	5156
456	928	4918	4763	7267	798	9737	7360
787	466	7194	7163	2848	71	67	2661
	2484	77	9650	9192	700	9280	4099
		21792	3083	36528	206	427	9730
					3618	24587	9222
							38861

Score on Part 5 = Number right × 10 = ...
[Total possible score = 70 points]

Copyright, 1925, by Scott, Foresman and Company

Skills Involved in Above Long Division Problem

Multiplication	Addition	Subtraction	Estimating Quotients
3×5	18+1	225	
3×6		−195	65)225
4×5	24+2	———	
4×6		305	65)305
7×5	42+3	−260	
7×6		———	65)455
		455	
		−455	
		———	

III. Summary

This chapter gives the reader a simple statement of the meaning and kinds of educational tests. The essential characteristics of educational tests are pointed out. A statement of what tests really do to serve the teacher is balanced by a few suggestions of things that tests do not and cannot do. This is followed by a discussion of the classifications of tests in general into the several possible categories. These are distinguished from scales; educational tests are distinguished from mental tests; individual examinations are distinguished from group examinations, and the limitations and the possibilities of each are pointed out. Achievement tests are discussed from the several possible classifications of rate tests, power tests, quality scales, product scales, survey tests and diagnostic tests. The accompanying chart should aid in clarifying the matter of the classification of the different types of tests. Each general group is illustrated by two representative tests in more or less common use. In the discussion of achievement tests particular emphasis is laid upon the importance of detailed and reliable diagnostic tests which afford the only sound basis for effective remedial teaching.

Chapter III presents a discussion of the uses which may be made of educational achievement and mental tests by the classroom teacher.

Classification of Educational Measuring Devices [23]

Tests						Scales	
Achievement			Mental			Quality	Product
General Survey		Diagnostic	General		Special		
Rate	Power		Individual	Group	Prognostic and Aptitude		
Courtis Series B Arithmetic	Thorndike-McCall Silent Reading Scales	Compass Diagnostic Arithmetic Tests	Stanford Revision of Binet	Otis Intelligence Examinations	Iowa Placement Examinations	Ayres Handwriting Scales	Iowa Spelling Scales
Burgess Silent Reading Scale	Stanford Achievement Tests	Charters Diagnostic Language Tests	Detroit Kindergarten Test	Terman Group Tests of Mental Ability	Stenquist Mechanical Aptitude Tests	Willing Composition Scale	Hahn-Lackey Geography Scale

[23] This is not intended as a complete classification of all educational tests. The tests listed are given merely as representative of the group in which they are placed.

SELECTED REFERENCES

Gilliland, A. R., and Jordan, R. H., "Educational Measurements and the Classroom Teacher." The Century Company, New York, 1925.

Gregory, C. A., "Fundamentals of Educational Measurements." Houghton Mifflin Company, Boston, 1923.

Kelley, T. L., "Interpretation of Educational Measurements." World Book Company, Yonkers-on-Hudson, New York, 1927.

McCall, W. A., "How to Measure in Education." The Macmillan Company, New York, 1923.

Monroe, W. S., De Voss, J. C., and Kelly, F. J., "Educational Tests and Measurements, Revised." Houghton Mifflin Company, Boston, 1924.

Ruch, G. M., and Stoddard, G. D., "Tests and Measurements in High School Instruction." World Book Company, Yonkers-on-Hudson, New York, 1927.

Trabue, M. R., "Measuring Results in Education." American Book Company, New York, 1924.

CHAPTER III

THE USES OF EDUCATIONAL TESTS

This chapter has for its purpose the discussion of the following uses of educational achievement and mental tests:

a. **The use of achievement tests for the purpose of comparing pupil with pupil, class with class, and school with school.**
b. **The use of achievement tests for the comparison of pupils, classes, and schools with norms and standards.**
c. **The use of achievement tests for determining the efficiency of instruction.**
d. **The use of achievement tests for purposes of class and individual diagnosis.**
e. **The use of achievement tests for classification, gradation, and pupil guidance.**
f. **The use of intelligence tests for classification, gradation, and guidance.**
g. **The inter-relation of achievement and intelligence test results.**

18. Tests in Relation to Classroom Instruction. So much comment has been made about the value of educational tests as teaching and supervisory instruments, as aids to school surveys, as research equipment, and as tools for administrative purposes, that very often the classroom teacher overlooks their real value in the solution of his own instructional problems. Yet this is where the most vital and important uses of tests are to be found. The development of reliable, valid, and highly detailed measuring instruments for the teacher has caused him to shift considerably his previous conceptions of the uses of tests. Earlier experience with tests tended to build

up in the teacher's mind the notion that tests were merely time-consuming devices used for checking up on teaching efficiency, and from which he at no time received any constructive help in the improvement of his instruction. The more modern use of tests implies their continuous use parallel with instructional progress. Upon the completion of an instructional unit the teacher will call to his aid a section of a detailed diagnostic test as a means of determining for himself whether his emphasis has been properly placed and the desired habits or skills properly developed. In a sense this means a continuous testing program, but experience with the other conception of the use of tests indicates that only through some such method will standard tests ever come to function at their highest efficiency as instructional instruments in the classroom.

I. Uses of Achievement Tests

19. For Purposes of Comparison. Since the earliest beginnings of group instruction classroom teachers have wished to know just how their pupils have compared in attainment with other similar pupils and classes. Until after standard tests were developed it was practically impossible to secure this information. Now the giving of a standard test in arithmetic, spelling, reading, or other school subject makes fairly easy a comparison of the results from a class with the norms and standards established for that subject and grade. For example, the use of the *Haggerty Reading Examinations* by a teacher furnishes the following data for comparison:

1. Individual pupil reading scores
2. Individual pupil reading ages
3. Individual pupil reading quotients
4. Average or typical reading scores for the class
5. Average reading ages for the class
6. Average reading quotients for the class

In this way a teacher may discover what should be expected

of a class as well as how well the class is progressing in a particular school subject.

Comparisons with other classes within the system in which the teacher is working, within the same building and even between different sections of classes in charge of the same or different teachers can be made on a basis of objective standards that have been derived for the various tests. Another sort of comparison which is often useful is that between the attainment of a class at the beginning and the end of a semester's or year's work, or at shorter intervals in the course of a semester. Each of these comparisons has its own peculiar value in assisting the teacher to determine the relative attainment and progress of his class at a given time. Such comparisons also serve to stimulate the individual pupil to greater effort and place upon him a more definite feeling of responsibility for his own development.

20. For Measuring the Efficiency of Instruction. Such general comparisons as are cited above are of great value in themselves, by way of indicating whether classes are up to standard in their achievements. But it is equally important to be able to determine as a result of these comparative studies whether the act of teaching itself may not be improved. Ambitious teachers everywhere are looking for the best methods of instruction in the different fields. Teaching methods can in the last analysis be evaluated only in the classroom by the classroom teacher. Standard tests offer the most effective instruments for such evaluation. Instructional units within the course of study must also be evaluated. The test of the effect of certain types of drill exercises, the determination of the specific strengths or weaknesses of groups or classes constitute uncounted opportunities for the use of these valuable instructional devices. Only through the use of large quantities of such available material in connection with numerous such simple yet significant investigations by the classroom teacher will the multitude of now unanswerable questions in this field be answered.

21349

21. For Purposes of Class Diagnosis. Very often at the beginning of a school term a teacher wishes to obtain advance information concerning the proficiency of his classes in certain subjects, and their general preparation for the work. He desires to know their weaknesses and their strengths, in order to direct his work in such a manner that the best results will be obtained. He desires to know the background his pupils have been given for the work they will be expected to master during the ensuing year. Many standard tests are now available which permit this type of use. It is not always necessary for a teacher to employ a special diagnostic test to find out all he desires to know regarding the relative abilities of the class. The ability to read silently is the basis of proficiency in so many subjects, that he should certainly test the ability of the class in reading. The results should indicate whether the class as a whole or the individual members of the class are able to interpret the printed page with facility, and so carry on their work without great assistance. It is also possible to test in a like manner for reasoning, judgment, and for other general qualities, as well as for knowledge of specific subject matter.

Not only is this preliminary diagnosis of great value to the teacher, but it has also been found desirable and valuable to check progress or advancement from time to time by means of objective tests. Tests of achievement will reveal whether the class as a group is moving together, or whether there are more or less well-defined sub-groups which seem to need special treatment, and which will furnish a justifiable basis for dividing a grade or class into sections for such special treatment. Where classes are divided into several sections as is often done in larger schools, the pupils should be arranged so that groups of approximately equal ability are placed together. The objective test is the best means for making this adjustment so that pupils can move forward in groups of nearly equal proficiency.

An excellent illustration of this need is given in Table 1 showing the rather startling range of mental ability found

in a typical 9th grade class. A teacher confronted with a class ranging in ability from above twelfth grade ability to below seventh grade is faced by a hopeless task if he is expected to bring all members of this class up to the same level of proficiency. Particularly is this true when the range of ability is wholly unsuspected as it is in so many cases. The systematic use of tests for class diagnosis similar to this constitutes a real source of professional protection to the classroom teacher.

Table 1

DISTRIBUTION OF MENTAL ABILITY OF A 9TH GRADE CLASS

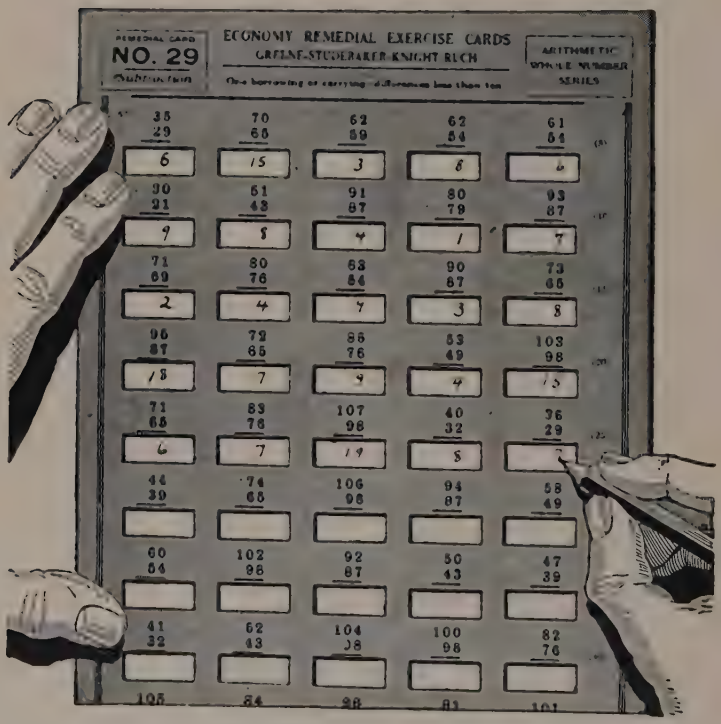
(According to Grade Location of Ability on Terman Group Test)

Grade Location	Number of Pupils
Above 12	6
12	4
11	12
10	8
9	19
8	8
7	6
Below 7	4
Total Number of Cases	67

22. For Purposes of Individual Pupil Diagnosis. Closely connected with the use of the test for diagnosis of a class or grade is its use for the determination of the difficulties and variabilities of each individual pupil. While, in general, individual differences are not so marked as to preclude efficient class instruction yet the more that is known about each child's weaknesses and strengths, the greater are the possibilities for success on the part of the teacher instructing the group. The test results should be studied especially in the light of each pupil's individual attainments, and points of

difficulty. The critical analysis of each pupil's test score may very likely be a means of clearing up wholly unsuspected troubles which would otherwise continue to hamper the child and to reduce his chances for proper advancement.

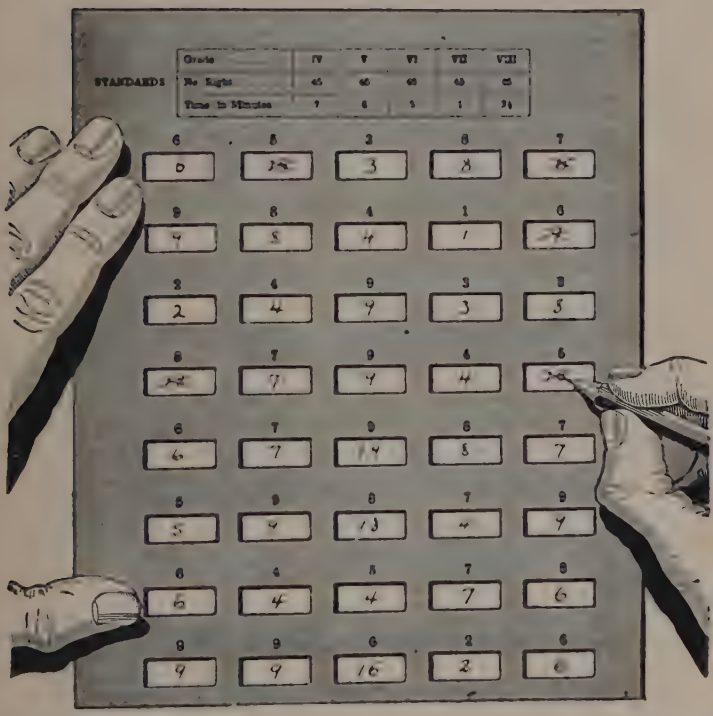
Specimen of Remedial Drill in Arithmetic



This type of individual diagnosis is especially valuable when considered in the light of various sorts of drill material so devised that each child may be aided in correcting his own weaknesses by drilling on his particular difficulties independently of the class. Perhaps one of the most significant developments of this type of material is the series of *Economy Remedial Exercise Cards* which have been constructed to

parallel the *Compass Diagnostic Tests in Arithmetic.*[1] The distinctive features of this material are clearly shown in the accompanying reproductions of the exercise cards in actual use.[2] These exercise cards are constructed in a specially

Specimen of Remedial Drill in Arithmetic



designed form which permits their repeated use.[3] No work

[1] "Economy Remedial Exercises in Arithmetic." Devised by Greene, H. A., Studebaker, J. W., Knight, F. B., and Ruch, G. M. Published by Scott, Foresman and Company, Chicago, Illinois.

[2] These figures are reproduced here with the permission of the publishers.

[3] See Chapter IX, pages 242 and 249 for further illustrations of this type of material.

is done on the cards themselves, but all work appears on the blank sheet of paper which is placed beneath the card while the pupil is at work. The exercises are so prepared that upon the completion of each unit of work in arithmetic a diagnosis of each specific difficulty of each individual pupil can be obtained by means of the diagnostic test, after which the remedial cards needed for the overcoming of these difficulties may be applied.

It must be clear from the foregoing discussion that one of the main purposes in the use of tests is achieved when precise and accurate diagnosis of individual pupil weakness is followed promptly by valid remedial drill in a manner such as is described above. Similar drill materials in a number of other subject matter fields are described in later chapters in this book.

23. For Pupil Gradation and Guidance. Teachers and supervisors find the problem of pupil placement one of the most difficult situations which they have to face. The indefinite lines of division between the grades and the wide overlapping of ability within the grades reveal the fact that our technique of pupil classification has been extremely crude. Practically all investigations have shown the typical grading of pupils to be unsatisfactory. Through the development of reliable grade and age norms, based upon the achievements of typical unselected school children in certain basic school subjects in terms of standard test scores, a valuable instrument for the improvement of this situation is placed in the hands of teachers, supervisors, and administrators.

By means of a simple procedure results from several standardized tests, even though they may express their results in different units, can be combined into a simple record and used as a valuable aid in pupil classification. The technique is as simple and as valuable regardless of whether a complete re-classification of a school or grade is planned or merely the proper placement of a few new pupils entering the system

for the first time is undertaken. By changing the test score into its equivalent grade-level [4] value a simple uniform graphic record of each pupil may be made. The following illustration will make this clear.

A pupil 11 years and 9 months of age presents himself for entrance into a certain school system. An examination of his credentials shows that the work he has been doing in the school from which he has transferred does not parallel the work in this school system sufficiently to make his classification a simple matter. The administration of four or five well-standardized tests in achievement together with one or two reliable group mental tests will provide data of much value in placing this child in the new school system at a level at which he will be able to do his best work and which will be thoroughly fair to him. Assume that one mental test and an achievement test in reading, arithmetic, language, writing, and spelling are used, and the following scores are earned:

Pupil: Robert K.	
Test	Score
Mental Test	118
Reading	38
Arithmetic	25
Language	21
Writing (Quality)	57
Spelling	61

The grade norms for the tests are as follows: [5]

[4] See Chapter VII, pages 195–198 and also the *Glossary*, page 379 of *Appendix* for explanation of this term and its use.

[5] These norms are hypothetical throughout, and are merely used for illustration.

Grade	Mental Test	Reading	Arithmetic	Language	Writing	Spelling
4	81	.....	13	.....	50	32
5	96	23	21	10	55	47
6	109	32	28	16	59	57
7	120	40	32	21	64	66
8	129	46	35	25	70	74
9	136	50	36			

By referring the score made on any of these tests to the table of grade norms it is possible to determine the grade level of achievement nearest which the accomplishment of this child places him. For example, a score of 38 on the Reading Test places him three-fourths of the way between the norms for the sixth and seventh grades. Accordingly, this score maybe thought of as 6.7 on the grade level scale. In a similar way the arithmetic score is four-sevenths of the distance between the norms for the fifth and sixth grades and this score may be considered as 5.6 points on this scale. The following graphic record card (Figure 1), almost universally useful in this field, is given merely as an example of how data of this type may be recorded and used. The critical and ambitious teacher will readily see that a purposeful use of tests such as is represented by this record card will reveal information about the pupil which is indispensable to his instructor.

The use of the grade-level method of commensurating test scores is of course only one of several possible methods. The chief defense for its use is its simplicity. Persons interested in a more careful classification of pupils on the basis of ability to learn as well as to make progress through the school will use the mental ages and achievement ages of their pupils for the purpose. Where it becomes necessary further to divide a grade or class once graded or classified by means of their mental age scores into two or more sections for greater convenience in instruction it is probable that the brighter pupils should be placed in one section while the duller

Graphic Record Card

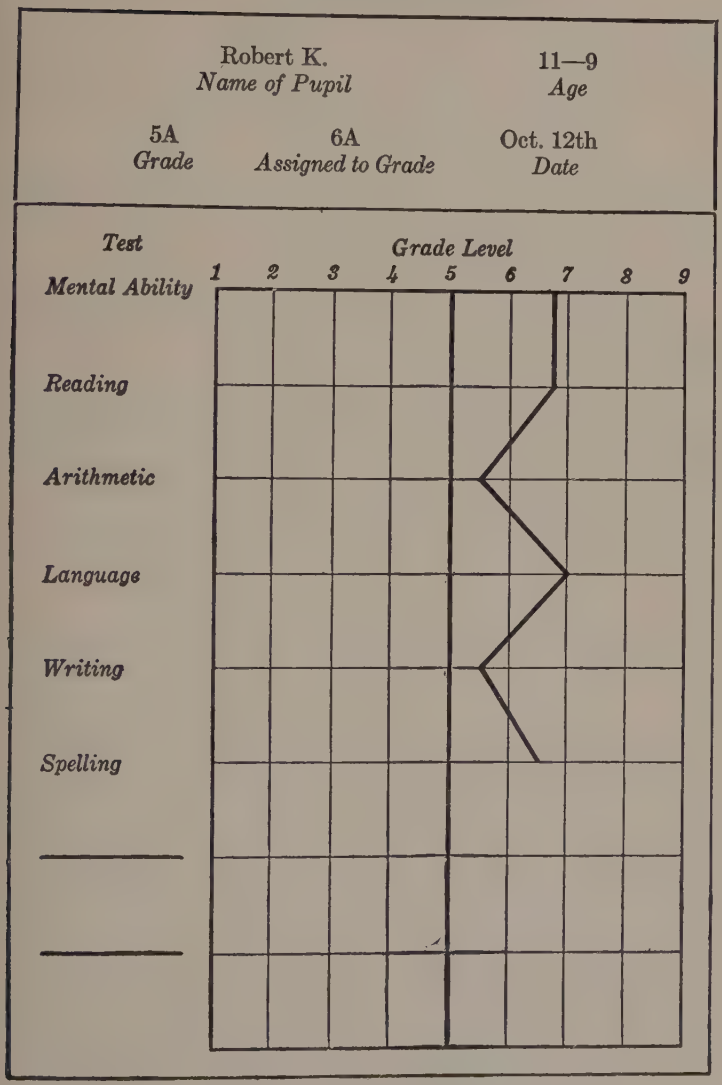


FIGURE 1. Pupil's Graphic Record.

pupils are kept in another section. In this case the intelligence quotient, which is an index figure expressing the ratio of the mental age of the individual to his chronological age, should be used.

Schools and school teachers are constantly being subjected to criticism because of the apparent waste from their failure to identify the special abilities of their pupils and to challenge these children to greater efforts. With the objective devices now available for the measurement of general as well as specific abilities this need not be. Teachers, principals, and college administrators have found that test records as such are extremely valuable aids in the handling of disciplinary cases and in many other ways. The majority of such administrative problems as disciplinary cases arise through the failure of the school system properly to stimulate and occupy the student's mind. Many of the reasons for such difficulties may be made clear to the teacher by the wise use of tests. The necessary adjustments can then be made to correct a situation which need not exist if properly handled.

II. Uses of Mental Tests

The meaning of mental or intelligence tests in general has been discussed in a preceding chapter (see pages 19–23). It remains to consider intelligence tests from the standpoint of their practical use to the classroom teacher. A large part of the rapidly growing popularity of intelligence tests among teachers and supervisors may be traced to three main causes; namely, (1) the tests themselves have been greatly improved in the accuracy and value of the resulting measures; (2) a larger proportion of school officers have become more intimately acquainted with intelligence tests and testing procedure with a correspondingly greater realization of the functions they serve; and (3) the changes in modern conceptions of education and attitudes toward it have made the utilization of such devices almost imperative. The stimulus given mental testing as a result of the gigantic army testing program dur-

ing the recent War resulted in the production of many types and forms of individual and group intelligence and general ability tests. Some of these have stood the wear and tear of time and some have been superseded by other more carefully constructed or more modern devices.

24. A Warning. The need for information such as is provided by any competent mental test is so great that no classroom teacher can be assured that her procedure with a given pupil is sound unless she has access to such exact information. The majority of group intelligence examinations must be used carefully and interpreted cautiously. Most of such group tests measure with sufficient accuracy for group comparisons but thus far no group intelligence test should be taken too seriously when it comes to individual comparisons. Many group mental tests permit the determination of mental ages (*M. A.*) and *I. Q.'s* (Intelligence Quotients) from the test scores. These derived scores are undoubtedly too unreliable to be taken seriously for any single individual. If it is important that the mental age or the *I. Q.* of an individual pupil be obtained and only group tests can be used not less than two forms of the test should be given at different periods. Even this procedure may result in conclusions which in many cases may be seriously erroneous.

On the whole, intelligence tests seem secure in the place they now hold as indispensable supporting agencies for achievement tests, as valuable instruments for the more exact classification and gradation of pupils in the classroom, as real guides to the teacher in matters of behavior and conduct as well as to the child himself in certain matters of vocation. There are, however, a few dangers attached to their indiscriminate use which the careful teacher and administrator should guard against. The more important of these are probably social in their character. In the first place, there is the danger which may arise through giving publicity to the results of intelligence testing. There is probably no other place in the use of educational measurements where so much

real "social dynamite" is to be found. The parents of children are reasonably tolerant of such devices until their own children are measured. If their children are tested in arithmetic and show up poorly, it is a matter of small consequence for that is the teacher's fault. It is a mere matter of instruction and reflects only on the school. But let one of their children show up poorly on a mental test and it is a different matter. Capacity is inherited and the results then do reflect on the parents. In the long run nothing but damage will be done by using intelligence test results for any other than purely administrative purposes and these in the strictest confidence.

A second danger in the careless use of intelligence tests lies in the effect which knowledge of such matters may have on the subject himself. The fact that a child is publicly labelled as unusually brilliant frequently subjects him to the dangers of unloosing a flood of egotism on his part, or depending upon his personality, subjecting him to the gibes of his fellow students. The fact that a child is publicly labelled as slow mentally may place a stigma on him which can never be removed. The only safe practice is to refuse consistently to divulge the results of such examinations to anyone not in a responsible administrative position and then to insist strictly upon their confidential use. Absolute adherence to this rule is the only safeguard. Precedents once established are hard to break. The value of the tests to the child and to the teacher is seriously reduced by an injudicious use of the results.

25. For Class Diagnosis. Viewed from the standpoint of the classroom teacher, achievement tests and intelligence tests are supplementary devices. After the teacher has given achievement tests and compared his class with the norms in a given subject he is still in danger of making several false assumptions relative to the significance of these results unless he has available further information such as is furnished by

intelligence tests. He may credit himself with an excellent job of teaching, when the innate brilliance of the group in his charge is such that if given really adequate instruction much superior results might have been achieved. It may be that the class falls considerably below the expected norm, and the teacher may feel that his teaching has proven a failure. This may also be an unwarranted assumption, for the class may be considerably below average in intelligence and can not be expected to approximate the norms that are set up for a class of average ability. It is evident then that there is a need for some means of determining approximately the intellectual capacity of the class. Mental tests have been devised to meet this need, and by giving several tests suitable for use in testing groups of pupils, the teacher can determine with a fair degree of accuracy whether his classes are up to the normal expectation in ability to master school work. The teacher will undoubtedly encounter a greater number of technical difficulties in both giving and interpreting mental tests than he finds in using achievement tests, but with practice and with the proper precautions he may feel that his conclusions are approximately correct for the class as a whole. He may, by comparing the results of several of these group mental tests, make a diagnosis of the ability of his class which will go a long way toward enabling him to determine what degree of proficiency he should reasonably expect in connection with the teaching of the regular subject matter. This knowledge will enable him to plan his work more intelligently, and to lay plans for overcoming difficulties and barriers which might otherwise have been unsuspected or delayed in making its appearance. Usually intelligence testing is done by the supervisory force, and the teacher simply calls upon the proper authorities for this item of information. While achievement testing may be thought of as essentially a part of the work of the classroom teacher, mental testing is more likely to be considered the work of the specialist, especially in the individual mental examination.

26. For Individual Diagnosis. The mental test may prove especially valuable to the classroom teacher in assisting him to solve the problems relating to the unusual child. The pupil may be unusually bright, troublesome, dull, mischievous, or in some other way quite out of the ordinary. The teacher may wish to know whether this child's typical responses actually reveal his real general ability; and whether or not the judgments of his former teachers and supervisors are correct. Mental tests will give information not obtainable in any other way. The results may indicate that the child has ability hitherto unsuspected; or that his supposed brightness is but a superficial pertness covering a really dull intellect. If it is a case of misconduct it may show that the child's bad behavior comes from the failure of the teacher to keep busy a really brilliant intellect, thus allowing his superfluous mental activity to find its outlets in mischief and disorder.

The mental test, when given to an entire group, frequently uncovers a child of outstanding capacity who has been content to go on with the group without especially revealing his real ability. Many such cases would never have been revealed, had it not been for the mental test. Mental tests reveal cases of overlapping in ability just as achievement tests reveal cases of overlapping in school achievement. They may reveal children in the fourth and fifth grades with the intelligence of normal seventh or eighth grade pupils, and may also reveal fourth and fifth grade pupils who in intelligence should be in the second or third grade.

When children are discovered who are far in advance of their place in school, readjustments of work should be made to match their abilities. This may be done in some cases by advancing them to a grade of work where their intelligence is given a real test, or by placing them in rapidly moving classes, so that they may progress according to ability rather than by some fixed promotion scheme; or by declaring for the entire class minimal requirements in the way of the different units of work to be done and then expecting the brighter pupils to attack the various problems at deeper

levels and more intensively than could be expected of the class as a group.

By proceeding along the same lines the dull pupil will also be better understood, because his difficulties can then be diagnosed more particularly and his strong points brought into relief, resulting quite frequently in an entire re-direction and reorganization of his instruction. In any case, the mental test will assist both in explaining difficult cases, and in revealing unsuspected general strengths and weaknesses.

27. Inter-dependence of Achievement and Mental Test Results. The teacher obviously should not attach too much importance to the results of intelligence tests or achievement tests alone. *Mental tests should supplement and not supersede the results of measurement of classroom achievement.* They should not be used in place of all other obtainable information about a pupil or a class, but should be interpreted in the light of such information. Accordingly, in order that too great stress may not be laid upon the child's proficiency in any one sort of test, the well informed supervisor or teacher attempts to rate the pupil in terms of his various attainments. By the intelligence test, the teacher discovers his mental "age," that is, how his general ability to learn agrees with a norm worked out for presumably normal children of any given chronological age. The teacher may find that a child 10 years and 6 months old, actually tests on the mental test as high as the "normal" child of 12 years and 3 months. This child is then said to have a "mental age" of 12 years and 3 months. His *I. Q.* (Intelligence Quotient) is the ratio of his mental age to his chronological age. In this case it is the ratio of 12 years 3 months to 10 years 6 months. Reducing to months, the ratio becomes 147 to 126 or 1.16. In actual usage it is the practice to multiply this figure by 100, expressing the *I. Q.* as the whole number 116. This means that this child is considerably accelerated in his mental development.

But let us assume the teacher has also tested the child in

other ways. His ability on a *Compass Diagnostic Arithmetic Test,* a *Barr Diagnostic History Test,* the *Iowa Spelling List,* or on a battery of achievement tests such as the *Stanford Achievement Tests* has been measured. The teacher here also finds his achievement age in these subjects, as compared to his actual chronological age, and works out ratios for these subjects, as between his chronological (real) and mental ages.[6] Although usually the norms of achievement tests are given in terms of school grades, and not in years and months, yet by using the ages usually assigned as average ages for the school grades, a reasonably accurate index may be worked out. Thus when a norm is given for a history test for the sixth grade, it may within reasonable limits of error be utilized as the norm for the average age of unselected sixth grade children.

The achievement age of the child may then be compared with his real or life age. The resulting quotient is usually called the *E. Q.* or the "Educational Quotient." If the *E. Q.* is then divided by the *I. Q.* of the pupil, the result is called the *A. Q.* or the "Achievement Quotient."[7] Franzen[8] has suggested that the accomplishment quotient be calculated for a combination of achievement tests instead of being based upon a single test. Buckingham and Monroe[9] have devised a set of combined educational-mental tests arranged so as to facilitate the calculation of the *A. Q.* Tests of this type give,

[6] McCall in his "How to Measure in Education" presents a table based upon the investigations by Ayres, Terman, and Kelly which gives the average age of a first grade child as 80 months, at the entrance into school, and adds 13 months to this figure for each succeeding year, as the investigators quoted seemed to agree that the average time spent in each grade is from 12–13 months. On this basis the average age of a sixth grade pupil is 145 months, or if the norm is for the month of May instead of September, as will often happen, then the age will be 154 months.

[7] See discussion on pages 203–205 of Chapter VII.

[8] Franzen, R., "The Accomplishment Quotient," *Teachers College Record,* November, 1920.

[9] *Illinois Examinations,* Public School Publishing Company, Bloomington, Illinois.

of course, rather rough measures, but have been extensively used for solving the problems of group diagnosis.

III. Summary

This chapter emphasizes the following main uses of educational tests by the classroom teacher:

1. Comparative Uses
   a. Pupil with pupil, class with class, school with school.
   b. Pupils, classes, and schools with norms and standards.
2. Instructional Uses
   a. Efficiency of teaching
   b. Evaluation of teaching methods
   c. Diagnosis of class difficulties
   d. Diagnosis of individual pupil difficulties
3. Pupil gradation and guidance
   a. Grade placement in accordance with ability
   b. Recognition of special aptitudes

In addition to this general treatment of educational tests, the uses of general mental ability tests are also discussed. Mental tests are important instruments for comparison, instructional, and guidance purposes. They are also recognized as fundamental to the correct interpretation of the meaning of achievement test results. The reader who is interested in making a critical examination of the uses which may be made of educational tests should study carefully the references at the end of this chapter.

The specific steps which should be taken by the teacher in the actual use of tests in the classroom are treated in Chapter IV.

SELECTED REFERENCES

Courtis, S. A., "Why Children Succeed." Courtis Standard Tests, Detroit, 1925.

Gilliland, A. R., and Jordan, R. H., "Educational Measurements and the Classroom Teacher." The Century Company, N. Y., 1925.

Gregory, C. A., "Fundamentals of Educational Measurements," Houghton Mifflin Company, Boston, 1923.

Kelley, T. L., "Interpretation of Educational Measurements." World Book Company, Yonkers-on-Hudson, New York, 1927.

Hines, H. C., "A Guide to Educational Measurements." Houghton Mifflin Company, Boston, 1923.

McCall, W. A., "How to Measure in Education." The Macmillan Company, New York, 1922.

McCall, W. A., "How to Experiment in Education." The Macmillan Company, New York, 1923.

Monroe, Walter S., "The Theory of Educational Measurements." Houghton Mifflin Company, Boston, 1923.

Monroe, W. S., DeVoss, James C., and Kelly, F. J., "Educational Tests and Measurements, Revised." Houghton Mifflin Company, Boston, 1924.

Otis, Arthur S., "Statistical Methods in Educational Measurements." World Book Company, Yonkers-on-Hudson, N. Y., 1925.

Pressey, S. L., and Pressey, L. C., "Introduction to the Use of Standard Tests." World Book Company, Yonkers-on-Hudson, New York, 1923.

Ruch, G. M., and Stoddard, G. D., "Tests and Measurements in High School Instruction." World Book Company, Yonkers-on-Hudson, New York, 1927.

Trabue, M. R., "Measuring Results in Education." American Book Company, New York, 1924.

CHAPTER IV

USING STANDARD TESTS IN THE CLASSROOM

The purpose of this chapter is the consideration of the following points involved in the effective and understanding use of standard tests in the classroom.

a. The classroom teacher's responsibility for the use of standard tests.
b. The factors which determine the time when tests should be given.
c. The responsibility of the classroom teacher for the scoring and interpretation of standard tests.
d. The specific steps in the inauguration of a testing program.
e. The importance of centering the testing program around some definite instructional or supervisory problem.
f. The criteria for the selection of tests for classroom use.
g. General rules for use in conducting a testing program.

28. Introduction. The value of the standard test to the classroom teacher is directly proportional to the extent that teaching practice is favorably affected by the results obtained. The problems involved in securing these results in the most effective and economical manner are the problems of the testing program.

I. Giving the Tests

29. The Teacher's Responsibility. In the earlier stages of development of standard tests it was believed that the most valuable results came from their use in a periodical survey by persons other than the classroom teacher. Recently it has come to be more generally accepted that as many of the tests

as possible should be given by the classroom teacher. This seems to be especially true in the case of tests which furnish information of special importance in the improvement of classroom instruction. In addition to allowing the classroom teacher to become acquainted with the technique of testing it gives him a first hand opportunity to observe the reactions of his pupils in the various test situations. On this account it is believed that wherever test results are to be used definitely as a basis for the discovery of individual pupil difficulties they should as far as possible be administered and interpreted by the classroom teacher himself. However, in cases where the test results are used for the more general purpose of a survey of achievement in the entire school or system, it is believed that it is less important for the teacher to have an intimate contact with the testing program. As a matter of fact many school administrators prefer not to have the teachers give the tests when they are used for such survey purposes. Furthermore in the case of certain types of tests such as the individual mental examinations which require special training for their administration, it is probably wise to discourage the average classroom teacher from attempting their use.

30. Scoring and Interpreting Tests. The matter of the responsibility for scoring and interpreting the test results constantly arises as an administrative problem in the smaller school. Teachers are likely to feel that the responsibility for scoring standard tests given for supervisory purposes should not fall to them. A part of this difficulty arises through a failure on the part of the administrators to make perfectly clear to the members of the teaching staff their responsibility with regard to this type of work at the beginning of their term of service. Most teachers, if given a suitable amount of time, do not seriously object to scoring test papers for their classes, particularly when they come to realize that this work will reveal information which will be extremely significant to them in the improvement of their

teaching practice. Teachers should feel that it is an opportunity rather than an additional burden or responsibility to be allowed to correct the test papers for their classes, and to study the results revealed by the tests. If a real interest in the outcome of the testing program is stimulated by the supervisory officers there will be little difficulty in inducing the teachers to help in the scoring and interpreting of the test papers from their classes.

II. When to Give the Tests

31. Periodical vs. Continuous Use of Tests. Tests may be used periodically or they may be used somewhat constantly in the classroom. The best adaptation of periodic vs. continuous testing depends somewhat upon the purposes the tests are to serve and the nature of those selected. If the tests used are the more common survey tests of general achievement they are usually given early in the school term, and then again a few days before the end of the school term. This procedure permits the teacher to determine the improvement which his pupils have made during this period. If he is a wise teacher he will realize that a portion of this measured improvement is due to sheer maturity on the part of his pupils and that a portion of it is the result of his instruction. The periodical use of educational tests to measure class or individual pupil improvement is by far the most prevalent practice. Tests which are used definitely for survey purposes and are given only once during the school year are frequently given at or near the end of the year. This is probably one of the least important times for tests to be given, since the entire school year is gone and there is no opportunity for the teacher to attempt to do anything about the conditions revealed by the tests. If only a single cross-section of the school is taken this should undoubtedly come early enough in the school year to permit the teachers to profit from the findings.

A further refinement of the idea of using tests early in

the school year to secure information for the guidance of the teacher is found in their use immediately following the completion of the instruction on a particular unit of subject matter. This means that many tests must be devised, each measuring a rather narrow field. This, of course, is exactly what is being done in the development of the detailed diagnostic tests which are now proving so useful. By using these narrow-function tests immediately after the completion of the teaching of a specific unit of subject matter the teacher may become immediately informed of the special weaknesses of his class. The special inadequacies of his instruction are thus made clear, and he can proceed at once to set up a remedial program before the class has moved on to other subject matter. This procedure represents the use of tests as the basis for remedial work. The same type of information provided by the use of these numerous narrow-function tests which guides the teacher in setting up the remedial program is also useful in organizing instruction in the future to prevent the appearance of these recognized weaknesses. The procedure represents the use of tests as a basis for preventive work.

III. The Basis for the Testing Program

32. Clear-cut Teaching Problem as Basis. One of the most common errors made by teachers and supervisors is the inauguration of a testing program without the formulation of a clear-cut problem, the solution of which can be most advantageously reached through the use of standard tests. The more clearly defined this problem is the better, for the testing program will thus be more limited in extent and increased in intensity. The best constructive supervisory work will result from careful intensive cultivation of a limited field. If the work is undertaken in this way much time will be saved and one of the most common criticisms avoided; namely, that the time of the pupil and teacher is taken for the testing and nothing ever comes of it. The teacher and pupils have

a right to profit from a knowledge of the conditions revealed by the testing.

33. Steps in a Testing Program. A tentative outline for a testing program is presented here as a suggestion for the general organization of the work.

1. Formulate and state a clear-cut teaching or instructional problem in the solution of which data from standardized tests will prove valuable.
2. Select the best available tests for the purpose.
3. Determine the time and method of administration of the problem.
4. Train the staff to be employed in the administration of the problem so that uniformity of conditions may be guaranteed.
5. Administer the tests.
6. Score and tabulate the results.
7. Analyze and interpret the results.
8. Propose a program which will result in the elimination or improvement of the conditions revealed.

A further discussion of these steps and their elaboration in connection with the problem of developing and using a testing program constitutes the purpose of the next pages.

34. Illustrations of Problems. Problems suitable to form the basis for a testing program are to be found in almost all of the fields of education. Frequently these problems overlap. On the other hand, quite frequently one test or a series of closely allied tests will contribute to the solution of several problems. The problems listed below are classified in accordance with their interest to administrators, supervisors, and teachers. It is clear, of course, that the list is not exhaustive.

A. Problems of interest primarily to teachers and supervisors

1. The determination of how near the pupils and the classes are to the standards in the different subjects.
2. The determination of the progress of the class in the different school subjects over a given period of time.

3. The determination of whether different subjects and phases of subjects are being properly or unduly stressed, as indicated by relative accomplishment.
4. The diagnosis of defects of pupils in the various subjects or in particular phases of a subject.
5. The determination of whether the pupil or the class is working up to capacity. Pupils of greater mental capacity should be expected to accomplish the most in their school work.

B. Problems of interest primarily to administrative and supervisory officers

1. The division of classes into two or more sections according to ability. Either intelligence or educational achievement tests or both may be used for this purpose.
2. The selection of pupils for special classes such as classes for the exceptionally bright or exceptionally dull pupils or for pupils having special defects in certain subjects.
3. The determination of the efficiency of the school as a whole by comparison of obtained scores with standard scores and with scores made by other schools or grades.
4. The determination of whether the proper emphasis is given to all subjects or whether some subjects are overstressed.
5. The comparison of different methods of instruction or comparison of new methods with the methods of instruction.
6. The determination of the general intelligence or achievement level of a grade, a school, or of a system.
7. The measurement of the progress of a grade, a school, a system, for a semester, a year, or for any given period of time.
8. The determination of whether or not the grade, the school, or the system is achieving what can be fairly expected for achievement.
9. The determination of the misplacement of pupils in grades or sections.
10. The proper classification of new pupils entering the school system.

IV. The Criteria for Selecting Tests

In the discussion of the necessary steps in the development of a testing program in Section 32 it was suggested that the first step is the selection of suitable instructional or supervisory problems around which to organize the proposed use

of tests. The next step in planning a testing program is the selection of the tests to be used. In this connection there are certain criteria which the teacher or supervisor will do well to keep in mind.

35. Validity. Is the test *valid?* Does it really measure the qualities, abilities, and skills, that it is supposed to measure? For a test to possess validity it is necessary that the materials actually included in the test be of prime importance, that the questions sample widely among the essentials over which complete mastery on the part of the pupils can be reasonably expected, and which they are given an opportunity to master in the course of study.

Validity may be defined more specifically as an expression of the degree to which a test measures the qualities, abilities, skills, and informations which it is designed and supposed to measure. In a general way it may be expressed in terms of the extent to which the test sets up a situation calling into play the use of the skills or abilities which experienced observers consider fundamental to success in a given field. Such judgments are represented by the opinions of authors of textbooks and courses of study, experienced teachers, and recommendations of committees, and other qualified authorities. Validity may be expressed statistically in terms of the degree of relationship (correlation) found between the particular measuring device and some other similar device the usefulness of which has previously been established. An illustration of this type of validation is given in Problem 39 of the "Work-Book in Educational Measurements."[1] In this case the criterion used is the judgment of the teacher who has had extensive contact with the individual pupils.

The source from which material in a test is selected is one index to the validity of the test. For example, a spelling test composed of words selected from an unabridged dictionary by taking the first word found on each tenth page of the book would certainly not be a valid test for at least

[1] Greene, H. A., "Work-Book in Educational Measurements," 1928.

two reasons. In the first place, the words would be of unknown social usefulness. In many cases it would not matter whether the pupils could spell the words or not. In the second place, the test would be lacking in validity because the pupils taking it would doubtless have had no opportunity to learn to spell these words.

A useful illustration of the application of the method of validating test material in terms of judgment of qualified authorities is found in the development of the *Iowa Silent Reading Test, Advanced Examination.*[2] The following tabulation shows the unit skills contributing to the pupil's ability to read and to work with books, which are measured by these reading tests.

Test 1. Paragraph Comprehension
- Social Science Material
- Science Material
- Literary Material

Test 2. Word Comprehension
- Social Science Vocabulary
- Science Vocabulary
- Mathematics Vocabulary
- English Vocabulary

Test 3. Sentence Comprehension

Test 4. Sentence Organization

Test 5. Paragraph Organization
- Selection of Central Thought
- Outlining
- Organization of Paragraphs

Test 6. Location of Information
- Use of Index
- Selection of Key Words
- Alphabetizing

Test 7. Silent Reading Rate

[2] Jorgensen, A. N., and Greene, H. A., *Iowa Silent Reading Tests, Advanced Examination,* Form A. Also *Examiner's Manual* for same. Bureau of Educational Research and Service, University of Iowa, Iowa City.

Compare with this list of measurable skills the major objectives of reading instruction given in Chapter XI, pages 288 and 292. The case for the validity of this reading examination rests on the exactness with which these objectives of reading have been paralleled in the test parts comprising it. In this case the validity of the test is determined by the extent to which it measures the desirable skills in silent reading as recognized by specialists in this field. In the case of achievement tests based more generally upon information rather than skills, the validity of the test depends more largely upon the opportunity the pupil has had to master the information measured in the test. In such a situation the teacher himself is probably the best judge of the validity of the test, since he knows best what material he has taught the class.

The general concept of validity may be made clear if the reader will think of the test situation as a small section (sampling) of a larger situation in life. The more nearly the test situation reflects the actual situation in life the more valid it becomes as a measure. In the attempt to devise tests which would be valid for the selection of persons for different types of activities during the War the last and final test was an exact duplicate of a life situation. Truck drivers who were being tested for their ability to take a supply or hospital truck into a shell-ridden area under cover of darkness were tested by actually having them drive a truck over shell-wrecked roads at night without lights. In such a case the test situation was a life situation. The successful passing of such a test was proof of ability. Inferences as to ability were not necessary. The validity of the test was absolute. The direct application of this conception of validity to standardized or informal tests is not always possible but where it is possible it is a useful criterion.

There is a growing tendency among users of standardized tests to place the responsibility on the makers of the tests for the establishment of the facts concerning the validity and reliability of the test. In the examiner's manuals for

the better types of standardized tests complete statements of the validity and the reliability of the tests will be found. Persons interested in the critical examination of a test soon learn to look for these facts first.

36. Reliability. Is the test *reliable?* Does the test measure consistently those qualities which it does measure? In order to secure reliability in a test the exercises comprising it must be selected carefully and extensively from the field which it is supposed to measure. This is illustrated clearly in the case of the examination of car-loads of grain for the market. It is obviously impossible for the purchaser to examine minutely every bushel of corn or grain in the car. It is possible, however, for specimens of the grain to be taken from different places in the car and from different cars. This process is called sampling. On the basis of these samples the car of grain is graded as to quality and on the basis of this quality its price is determined.

In a similar manner exercises or items of information are sampled from the possible items suitable for a test in a given field. The accompanying diagram (Figure 2) illustrates the effect of sampling on the reliability of test exercises based on a certain limited field of information. Each of the 40 rectangular spaces in Figure 2 represents an item of information which an individual pupil (Pupil A) has had a chance to learn. The 28 shaded spaces represent the 70% of the items which he has actually mastered. The 12 unshaded portions are items which he has not learned. These items are thus 30% of the total. Let us assume that a very limited sampling at one end of the field of his information is made. If Exercises (items) 1 to 5 inclusive are selected he will fail on all but one, item 5. On the other hand if five exercises from the opposite end of the field of information are chosen he will fail on only one, item 18; thus a variation of from 20% to 80% on the items. From this it is clear that it makes a difference where the sampling is taken. This factor is largely taken care of, however, if the sampling is adequate

in extent. If the even numbered items are selected the pupil will fail on half of them. If the odd numbered exercises are

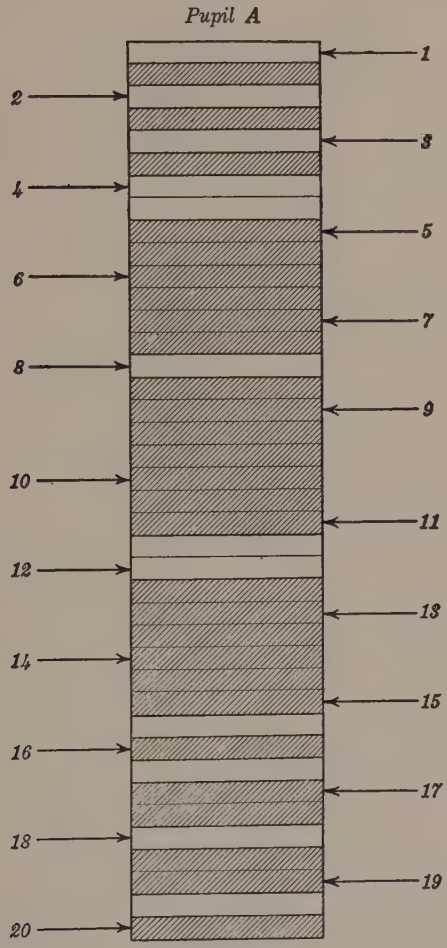


FIGURE 2. Illustrating Principle of Sampling.

chosen he will fail on only two. This results in a variation in his score of from 50% to 80% due largely to chance. If the twenty items are selected the pupil fails on 35% of the

exercises, and answers 65% correctly. It is to be noted that as the number of items selected is increased the pupil's success on the test more nearly approaches the actual amount of his information in this field. Thus it is clear that the extent of the sampling which the exercises in a test represent is an important factor in its reliability.

Increasing the reliability of tests means that chance factors such as temporary physical disturbances, fatigue, etc., which affect the consistency with which tests perform, must be more largely eliminated from the results. It means that much more extended testing periods and longer tests will be required. Serious doubts are beginning to be cast upon many general survey tests which limit their period of sampling to one period of two, five, or even ten minutes. Extremely limited and narrow skills may be sampled in that time but it is very doubtful if it is possible to sample such highly complex skills as many of our school subjects represent in that period of time. However, the amount of time required for giving and scoring a test should be considered with reference to the amount of information yielded. The time and money cost of tests may come to be calculated in terms of the cost per unit of information secured from its use.

Reliability, which is the index to the consistency of performance of a test, is commonly expressed in terms of the degree of correspondence between results from consecutive trials of the test, or between results from two forms of the same test. This procedure is illustrated concretely in the case of Problems 40 and 41 of the "Work-Book in Educational Measurements." [3] Since the process is somewhat technical and involves methods of correlation which have not been explained up to this point in this book, the student should probably postpone the examination of these illustrations until the statistical methods involved have been explained and mastered.

37. Objectivity. Is the test *objective?* Is it possible to administer it and score it without the results being affected

[3] *Ibid.*

by the teacher's personal judgment? Historically, objectivity is one of the first qualities of examinations to be noted by teachers. The need for the removal of the subjective factors in the marking of examination papers is one of the reasons back of the development of standardized tests. In fact, it is mainly because of the removal of the personal judgment factor by the form in which the exercises are stated, and the use of mechanical methods of scoring the exercises, that standardized tests are now considered to be superior to the traditional type of examination. Objectivity is an important requirement of a good test, but must be considered as secondary to validity and reliability, although it may contribute to each of these qualities indirectly. *Objectivity in a test exercise makes for the elimination of the opinion of the person who scores it.* This means greater accuracy in the grading of the exercises, and naturally greater reliability of measurement results.

In general, objective test exercises are stated in such a manner that only one answer satisfies the requirements of the statement. A common form of objective exercise is the recall type. Other types of objective exercises are the true-false, multiple response, matching exercises, etc. These and other forms of exercises are illustrated on pages 105–111 of this volume. The distinct advantage of selecting perfectly objective exercises for use in educational tests is that there can be a perfect agreement as to what is the correct answer. This means that outside of purely chance errors there should be no variation in the scores assigned to a given test paper by different persons.

The effect of objective exercises on the accuracy of marks assigned to a test or examination paper is illustrated in Table 2. This table shows the relative accuracy of marks assigned by a group of ten teachers to two types of test papers written by the same pupil on the same subject matter. Each teacher graded each paper independently and at different periods so that there was little chance for memory of the grade assigned previously to carry over. One of the

papers was a typical ten question essay-type examination in the field of Civics. The other paper was a 40 exercise true-false test based upon the same subject matter. It will be noted from Table 2 that the average mark assigned was 83,

Table 2

MARKS ASSIGNED BY TEN TEACHERS TO AN ESSAY-TYPE EXAMINATION IN CIVICS, AND TRUE-FALSE EXAMINATION OVER THE SAME MATERIAL

TEACHER	Mark Assigned	
	Essay-type (10 Questions)	True-false (40 Exercises)
A	84	31
B	80	32
C	88	31
D	76	30
E	82	31
F	90	31
G	85	31
H	82	31
I	83	31
J	80	31
Average.....	83	31
Average error (deviation from the average)...	3	.2

on the essay-type examination paper, and 31 points on the true-false objective test. It will also be noted that the highest mark assigned to the essay-type paper was 90 and the lowest was 76 points. All but two of the teachers agreed, however, on a mark of 31 as correct for the true-false examination.

The relative objectivity of the two types of examinations is shown definitely by the variation in the marks assigned by these ten teachers. In the essay-type examination each teacher missed the average score by an average error of 3 points. In the case of the objective examination paper the average error made was .2 point, and this error was contributed by two teachers, Teacher B and Teacher D.

Objectivity may be expressed statistically in terms of the degree of correspondence between scores or marks assigned to different papers by the same individual at different times. Problem 42 in the "Work-Book in Educational Measurements" [4] affords a concrete illustration of this method of determining objectivity.

38. Educational Significance. Are the abilities measured in a test being considered for use educationally significant? Do the fields measured by the test really represent *important and desirable outcomes in education?* Are they sufficiently definite to respond directly to instruction, or are they indefinite and intangible? An Organization Test devised a number of years ago by one of the authors of this volume is an illustration of a test which measures an ability or a composite of many abilities which it is extremely difficult to identify as an important and improvable outcome of school training. Organization ability is thought to be a factor in language ability and in silent reading ability, but so far very little is known about how to develop and improve it. This difficulty makes one somewhat skeptical of the educational significance of ability measured by the test.

39. Administrability. Is the test easily administered? This characteristic of a test is probably best expressed by the word "administrability." It represents the ease with which the test may be given, scored, and interpreted. Ease of administration must be considered from two distinct points of view. In the first place, the administration of the test

[4] *Ibid.*

from the standpoint of the pupil must be considered. Clear, simple, and direct instructions to the pupil should be printed on the test itself, and in most cases these directions should be given orally by the examiner. If some particular type of response is called for this should be illustrated clearly by one or more samples.[5] The second phase of the problem of administrability of a test is found in the directions to the examiner. Ordinarily the instruction to the examiner should not be printed on the test booklet. The instructions which accompany many of our tests are entirely too detailed to go on the test itself. Furthermore a large amount of information concerning the construction of the test, its validity, reliability, objectivity, the norms, and the methods of interpreting them, is usually necessary if the test is to be effectively used. Therefore it seems that for the most of the better and more useful tests a carefully written manual is almost indispensable.

Answer keys conveniently arranged to make for easy and accurate scoring of the test papers also contribute to the ease with which a test may be used. Class record and tabulation sheets which merely by means of the method of tabulation used enable the teacher or examiner to identify and locate the pupil who makes a given low or high score greatly add to the utility of the test. Such a class record sheet as that proposed for use with the *Pressey Diagnostic Tests in English Composition,* or that furnished with the *Courtis Silent Reading Tests, Series R*[6] will be found useful for this purpose.

The administrative usefulness of norms for tests is illustrated in Exercises 16, 17, and 18 in the *Appendix* of this book. Problems 30 to 38 in the "Work-Book in Educational Measurements" also give much concrete experience with this phase of test interpretation.

[5] Illustrations of excellent types of directions to students and satisfactory samples for certain types of objective test exercises are given on pages 105 to 111 of this book.

[6] See list of tests and publishers on pages 335 to 336 in *Appendix.*

40. Economy. Is the test reasonably economical? Economy is certainly not one of the major criteria to consider in selecting a test, for a cheap test may readily be useless. However, many users of educational tests are beginning to be forced to a consideration of this factor by the rapid increase in the numbers of tests available for use in different fields, and the corresponding increase in the per pupil cost of testing. Real economy will not result from the indiscriminate use of cheap tests. Cost of testing should probably in the long run be computed in terms of the reliability or validity of the test per unit cost. There is a very real possibility for effecting economy in testing without reducing the quality of the content of the test by the use of specially designed test booklets which permit the tests to be used repeatedly. The development of this type of test booklet is in the future, but in it lies the most hopeful possibility of reducing per pupil cost in testing without sacrificing anything in the quality of the test content.

41. Adequacy of Test in Measuring Specific Functions. In selecting a standard test for classroom use the six criteria previously noted for a good test should be kept in mind. Furthermore, a test should be selected because of its power to perform some specific function. If a general survey or superficial general diagnosis constitutes the purpose of the testing program the tests may be chosen accordingly. If a more detailed individual diagnosis is desired then tests adapted to that purpose must be selected. This may be specifically illustrated in the case of arithmetic.

Achievement in arithmetic is the result of many skills. A general diagnosis involves simply locating pupils who have and have not achieved the general objectives set for them. The *Courtis Standard Research Tests in Arithmetic, Series B,* are considered by some to be diagnostic in the sense that they measure achievement separately for each of the operations in the field of integers. They, of course, do not reveal the pupil's particular limitation. They simply indicate the ex-

tent to which he is lacking in the general ability to do certain types of examples in each of the four fundamental operations with whole numbers. Diagnosis of this vague and general sort is afforded by many other general survey tests.

To ascertain the particular types of example which he is unable to do satisfactorily it is necessary to use a more elaborate type of test such as the *Buswell-John Diagnostic Tests for Fundamental Processes in Arithmetic* or the *Compass Diagnostic Tests in Arithmetic.*[7] This is diagnosis proper. A general achievement test cannot reveal this detailed information. In most cases a study of a pupil's scores does not reveal why he failed to achieve or what particular errors he has made. In planning effective remedial instruction it is frequently helpful to secure this information by analyzing the pupil's performance. For an illustration of this type of analysis see pages 235–237. This is going beyond measurement in the sense implied by the ordinary use of educational tests. Such analytical diagnosis is particularly profitable in the case of pupils whose achievement is markedly below standard.

V. General Rules for a Testing Program

The following rules are common to most tests now in use. They do not take the place of the directions accompanying the various tests that may be used. The directions for giving and for scoring given in the Examiner's Manual which accompanies the better tests should be rigorously adhered to, so that the tests may be given under standard conditions. It is hoped that the following suggestions may be an aid in clarifying the administration of certain tests.

42. Preparation for Administering the Test. Any of the satisfactory educational tests can be given with entire success by any teacher or principal who is reasonably skillful in discipline, and who will carefully follow the directions accompanying the tests. Unless the test directions are extremely

[7] See list of tests and publishers on pages 335 to 336 in *Appendix.*

familiar the Examiner's Manual should be carefully studied before attempting to give the test. If possible the test should be tried on some other person in order to gain further familiarity with the procedure. If it is not possible to administer the test prior to its use in the classroom the directions should be read aloud several times so that they may be followed easily during the giving of the test. Familiarity with the directions is essential if the standard conditions for the test are to be maintained, and comparison with the norms and standards thus be made possible.

Pupils may be tested in ordinary classroom groups or in larger groups. In case several grades are to be given the same test, time may be saved by moving the combined group into a larger room, care being taken that the seats and the desks are adjusted to the size of the pupil. The probability is that the pupil will get a better score on the test if he is in familiar surroundings while taking it.

Before beginning the test, have the desks cleared and see that each pupil is provided with one or two sharpened pencils. Have at hand a number of extra pencils for emergencies during the examination. The room should be quiet throughout the test. Require strict obedience and attention. No questions should be allowed during the test. In giving directions speak distinctly and at a moderate speed. Cultivate a manner which is agreeable but which at the same time suggests authority. Pupils should be made to feel "at home" in taking the test. Pupils will look forward to taking the tests without fear or nervousness if they are properly given and no misconceptions with regard to the meaning and use of the results are allowed to arise.

Throughout the examination commands should be given in a wide-awake manner. The directions should be given slowly and with careful attention to emphasis. The voice should be just loud enough to carry to all parts of the room. The directions of the tests should be followed verbatim. As far as possible prevent disturbances within or without the room which might interfere with the administering of the test. To

avoid interruptions it is suggested that the teacher prepare a card carrying these words: *Testing Going On. Please Do Not Disturb.* If this card is hung on the outside of the classroom door interruptions will be less frequent.

43. Timing. The time limits as set in the directions for giving the tests should be strictly adhered to. This is especially true in short tests where a few minutes' or even a few seconds' error may invalidate the results. Tests should be timed to the second, or the results will not be comparable to what others get when the exact time is taken for the test. In timing the test a stop watch is almost essential. If an ordinary watch is used, it is necessary to exercise great care to secure accurate timing. In such cases it is well to have an assistant whose sole duty is to act as time keeper for the test. If this is impossible the following method will serve quite well:

	Hr.	*Min.*	*Sec.*
(*a*) Record time starting signal is given.....	11	18	20
(*b*) Add to this the time required for the test		5	00
(*c*) The sum is the time to signal a stop.....	11	23	20

44. Scoring. The scoring of most standard tests is made almost wholly objective by the use of scoring keys. The answer keys and directions for scoring each specific test should be followed rigorously. Score exactly as instructed even though it may seem contrary to your best judgment. This is very important in order that the scores may be compared directly with the norms or standards which have been derived for the tests. It is best that all necessary calculations be performed twice, and that all transcribed records be checked against the pupils' test papers to see that no errors have been made. Errors of this character may readily invalidate the entire testing program.

The balance of the testing program which follows the giving and scoring of the test is largely statistical in character and

on this account is discussed in Chapter VI of this volume, "Summarizing the Results of Testing," and in Chapter VII, "Interpreting the Results of Testing."[8]

VI. Summary

This chapter considers the responsibility of the classroom teacher in the giving, scoring, and interpretation of educational tests. In addition it points out some of the factors which determine when tests should be given, and the conditions under which a testing program should be inaugurated by a teacher or supervisor. Concrete suggestions of procedure in setting up a testing program are given along with a number of illustrations of problems suitable as a basis for testing programs. The major criteria of validity, reliability, objectivity, administrability, and economy, which should be considered in the selection of educational tests for classroom use are discussed and illustrated. In conclusion, a number of general rules for the conduct of a testing program are given. These suggestions will prove helpful to the inexperienced teacher and student, and will also assist in rendering the testing work of the experienced teacher and supervisor more effective.

This chapter emphasizes particularly standardized tests. The following chapter, Chapter V, considers the informal objective tests as developed, used, and interpreted by the classroom teacher.

[8] Practical problems in the application of the criteria for the selection of tests are given in the "Work-Book in Educational Measurements" which is designed as a companion volume for this book. The teacher or student who has had the necessary statistical preparation may turn at once to these problems (Problems 29 to 32) in the *Work-Book*. Teachers and students who have had no experience with this type of work should wait until they have mastered the earlier problems in the *Work-Book* before attempting these particular problems.

SELECTED REFERENCES

Abbot, A., "Tests for English Teachers." *English Journal,* 12: 663–71.

Briggs, T. H., "An Examination in First Term Latin." *Classical Weekly,* 16, No. 19, March 19, 1923.

Greene, H. A., and Jorgensen, A. N., *Examiner's Manual for Iowa Silent Reading Tests.* Bureau of Educational Research and Service, University of Iowa, Iowa City, 1927.

Kelly, F. J., "Teacher's Marks." Teachers College Contributions to Education, No. 66. Pages 1 to 84 and 133 to 134. Columbia University, New York, 1924.

McCall, W. A., "How to Measure in Education." Pages 119–133. The Macmillan Company, New York, 1922.

Mead, A. R., "Suggestions for the Training of Teachers in the Use of Educational Measurements." Educational Administration and Supervision, 12: 23–43, January, 1926.

Monroe, W. S.; DeVoss, J., and Kelly, F. J., "Educational Tests and Measurements (Revised)." Houghton Mifflin Company, Boston, 1924. Chapters X and XIII.

Monroe, W. S., "Measuring the Results of Teaching." Houghton Mifflin Company, Boston, 1923. Chapters XI and XII.

Ruch, G. M.; Knight, F. B.; Greene, H. A.; Studebaker, J. W., *Examiner's Manual for Compass Diagnostic Tests in Arithmetic.* Scott, Foresman & Company, Chicago, 1925.

Sears, J. B., "The School Survey." Houghton Mifflin Company, Boston, 1925. Part II, Chapter XIV.

Starch, Daniel, "Educational Psychology." The Macmillan Company, New York, 1919. Chapter XXII.

Toops, H. A., "Trade Tests in Education." Teachers College Contributions to Education, No. 115. Columbia University, New York, 1921. Pages 39–62.

Wood, Ben D., "Measurement in Higher Education." World Book Co., Yonkers-on-Hudson, New York, 1923.

Wood, Ben D., "Studies of Achievement Tests." *Journal of Educational Psychology,* Vol. 17, January, February, and March numbers, 1926.

CHAPTER V

CONSTRUCTING AND USING INFORMAL OBJECTIVE EXAMINATIONS IN THE CLASSROOM

The purpose of this chapter is the presentation of an introductory discussion of the following points which demand consideration when using informal objective tests in the classroom:

a. The ways in which informal objective tests and standard tests are alike in general appearance and use.
b. The major advantages and disadvantages of informal objective tests as substitutes for the essay or common type of examination.
c. General principles which should be understood and utilized in constructing informal objective tests.
d. The major types of objective tests.
e. The problem of "guessing" in objective examinations.

I. Characteristics of Informal Objective Examinations

45. Introduction. The history of the development of educational measurements indicates clearly the general dissatisfaction with traditional examination methods. Early studies such as those of Starch and Elliot, Johnson and Kelly, and more recent investigations such as those reported by Ruch [1] showed clearly the unreliability of all types of examinations in which the opinion or judgment of the examiner is a factor. The developments in the field of objective measurements have followed two main fields: (1) the development of standardized test material; and (2) the development of the informal

[1] Ruch, G. M., et al., "Objective Examination Methods in the Social Sciences." Scott, Foresman and Company, 1926.

objective type of classroom tests, frequently called the "new examination." Despite their similarity there seems to be little real conflict between these two types of test material. The meaning and the uses of standardized test materials are discussed in preceding chapters of this book.

46. The New Examination. One of the most frequently discussed modern educational topics is the so-called "new examination."[2] Even though standardized instruments for measuring achievement of school children have become very widely used, particularly in the elementary school, the examination set by the teacher still remains the most frequently used means of measuring the achievement of pupils. Although properly constructed standardized educational tests are superior in certain respects to examinations set by the teacher, they will probably never entirely displace them as a means of measuring the results of teaching as indicated by pupil attainment. The teacher frequently has need for a measuring instrument adapted to a particular course of study, or to the emphasis which has been given to a particular course of study, or to the emphasis which has been given to the subject in the teaching of a particular class. A standardized educational test must of necessity be adapted to generally recognized educational objectives, and to the typical or average course of study. Hence, such a test may be lacking in validity for a particular type of use and thus may not always meet the teacher's need. On the other hand, an examination may be constructed by the teacher to fit the instruction which the class has been receiving.

A comparison of the use of the essay-type examination question and the objective (true-false) type of exercise based on the same material is given in the accompanying illustration of an attempt on the part of a teacher to make the test

[2] The expression "new examination" is commonly used as a collective designation for unstandardized tests and exercises which may be objectively scored and which require little writing on the part of those being tested. They represent a definite movement away from the "essay-type" of examination.

exercises fit a particular item of instruction. The essay-type question designed to force the student to organize his thoughts on a unit of Civics material is as follows. A pupil's answer to this question, rated by a number of teachers at about 70% of a perfect answer, is also given.

Q. What are the purposes for which a political party is formed?

A. "The purposes for which a political party is formed are: (1) to establish a platform that will express the views of the members on certain vital questions; (2) to nominate a candidate or candidates for the purpose of running for office. They in turn will uphold the party platform; (3) to separate themselves from other parties with whose principles they do not agree; (4) to establish themselves as a separate organ with certain principles."

The following exercises show how a group of true-false statements based upon the same general subject matter as the above question may be used as an objective type test. It will be noticed that the pupil answered one of these four exercises incorrectly.

Directions to Pupil: Study each statement below. If a statement is True underline True; if False underline False. DO NOT GUESS.

1. A political party is formed by a group of people with common political views.	True	False
2. The purpose of a political party is to stimulate interest in political appointments.	True	False
3. Political parties usually cater to the opinions of the minority.	True	False
4. The actual accomplishment of political policies is dependent upon the election of those favoring the policies.	True	False

47. Informal vs. Standardized Tests. Standard educational tests are not fundamentally different from informal examinations in their basic elements. In both cases, the pupil responds to certain exercises. This performance is described by means of a score or "grade." Standard educational tests

are essentially nothing more than improved and refined examinations. Hence, the requirements of test construction suggest the criteria by which the written examination may be improved. It is obvious, of course, that in so doing it is not possible for the teacher to follow out in detail the complete procedure of constructing standard educational tests.

Many writers and speakers have advocated that objective tests [3] be employed either in conjunction with examinations of the traditional [4] type or in place of such examinations. Although many of those who have done so have based their opinions largely upon theoretical arguments, others have presented experimental data to support their judgments. Among the latter are Ruch,[5] Wood,[6] and Russell,[7] each of whom strongly recommends the use of objective exercises. Another experimenter,[8] however, has concluded that under satisfactory conditions there is little difference in the general merit of the two kinds of measuring instruments. The bulk of the evidence, however, seems to favor some use of the objective type of test.

[3] A test or measuring instrument is said to be *objective* when different persons using it obtain the same score or appraisal of the pupil's work, or the same person assigns the same score at different times. In other words, objective scoring eliminates differences due to the fact that the standards of different teachers and also those of the same teacher vary at different times.

[4] Term applied to examinations of the essay- or discussion-type which have been in common use for many years.

[5] Ruch, G. M., "The Improvement of the Written Examination." Scott, Foresman & Company, 1924, p. 193.

——— "An Investigation of Objective Examinations in the Social Studies," Scott, Foresman & Company, 1926.

[6] Wood, Ben D., "Measurement in Higher Education." World Book Company, 1923, p. 337.

[7] Russell, Charles, "Classroom Tests." Ginn & Company, 1926.

[8] Brinkley, Sterling G., "Values of New Types of Examinations in the High School." Teachers College Contributions to Education, No. 161, Columbia University, 1924, p. 121.

The results obtained by Brinkley are somewhat at variance with the investigations of almost all others. This is possibly due to the fact that his teachers were inexperienced in the use of new type examinations and skillful in the use of traditional examinations.

II. Advantages and Disadvantages of the Informal Objective Examination

48. Advantages of the Informal Objective Examination. The new type examination has the following significant advantages over the traditional type.

Objectivity: In an objective test the exercises are stated in such form that the answers are brief, and usually only one correct answer is possible. This makes it possible for the exercises to be scored or graded by different individuals without affecting the accuracy of the grading. A perfectly objective test may be scored repeatedly by one person with practically no variation in the scores, or it may be scored by a large number of persons with practically no variation in the scores assigned. Thus in the new type of examination the exercises may be graded on an impersonal basis, entirely independently of the personal judgment of the examiner. This is true, of course, only when the exercises are constructed in accordance with certain recognized principles which are discussed in a later section of this chapter.

Reliability: The use of objective or near-objective exercises in a test makes it probable that the reliability will be high. The reliability of a test is the degree to which successive applications of the test yield scores equivalent to those obtained from its first use. In other words, it is a measure of the variable errors in the test scores. Such factors as the disagreement of different persons as to the importance of a certain test exercise, or the correct answer to it obviously affect its reliability. The length of the test in terms of the time required to take it or in terms of the number of items comprising it is one of the significant factors affecting reliability. Reliability is discussed in more detail in Section 36, Chapter IV.

Brevity of Pupil Response: Because of the nature of the exercises the amount of writing done by pupils in responding to objective tests is reduced to a minimum. This practically

eliminates the advantage of rapid and fluent writers over those not so gifted. The fact that a student can write more material in the same length of time should not result in his receiving higher marks in his school subjects. Also there is much less opportunity for the teacher to be influenced, frequently unconsciously, by good or bad writing, spelling, punctuation and other similar qualities, which have little or no connection with achievement except in the field of English.

Extent of Sampling: The brief response by which a pupil is able to indicate his answer to an objective type exercise makes it possible for him to respond to a great many more exercises in a given time. Accordingly a much wider range of achievement can be measured or a much larger sampling can be secured than is possible by means of a discussion- or essay-type of examination requiring the same amount of time. This minimizes the likelihood that pupils just happen to know or not to know a large enough proportion of the answers to the questions asked that their scores are largely the result of chance. The operation of this factor of sampling is shown in Figure 2, on page 77. It is shown by this illustration that the typical discussion-type of examination with its few questions (limited sampling) may be a very unfair and unreliable measure of a pupil's information or ability especially if the teacher happened to use in his ten-question type of examination items number 1, 2, 4, and 18, when he might just as well have used items 5, 6, 11, and 19. Clearly the extensive sampling afforded by the use of the objective exercises represents a measure of protection for the student as well as the teacher from unfair and unreliable measurements.

Other minor advantages of the informal objective type of examination might be enumerated. It seems, however, that these four major items are sufficient to justify the use of objective exercises in preference to those of the traditional type when one desires to measure detailed factual knowledge. It must, of course, be admitted that there are certain subject matter fields in which the discussion type of examination

will necessarily remain the most useful measuring instrument. It is also undoubtedly true that even in the fields for which it is best suited (informational and content fields) it would be unwise to make the objective examination the sole basis of measurement.

49. Disadvantages of the Objective Examination. When the fundamental functions of the written examination are all considered there seems to be only one point at which the objective type operates at a serious disadvantage. As agencies for the stimulation of students to more effective effort there is little or no evidence of superiority of one type over the other. As agencies for the more effective measurement of pupil progress and mastery the evidence definitely favors the objective type. However, from the standpoint of one commonly accepted function of examinations, training in thought organization and written expression, the objective examination seems somewhat lacking. A further disadvantage which is highly related to this is the fact that many of the tests of the objective type appear to place an undue emphasis on the memorization of details with little or no development of discrimination, evaluation, or organization ability on the part of the pupil. In a sense this appears to be a serious objection, but more careful consideration reveals the fact that it is a function of the methods of constructing the examinations and is not innate in the examinations themselves.

III. Principles of Constructing and Using Objective Exercises

Before attempting to construct informal objective test exercises the student and classroom teacher who is inexperienced in this type of work should consider the following general principles underlying their construction and use.

50. Need for Specific Directions. *Objective test exercises should always be preceded by explicit directions as to just*

what is to be done. If the pupils are not familiar with the kind of exercises used in the examination the directions should also include samples showing just how to do the work. The amount of time to be used should also be stated in the directions if time limits are to be used. Observe this sample set of directions for a true-false examination.

True-False Examination

Directions to the student: Examine each statement below and decide whether it is true or false. If the statement is true, underline True; if the statement is false, underline False. DO NOT GUESS. This examination will be corrected for guessing. You will be given ——— minutes for this examination. Study the sample below. It has been answered correctly.

SAMPLE: The reliability of a test means the extent to which a test measures whatever it does measure. True False

51. Number of Exercises to Use. It is important that a fairly large number of items[9] be included. If such tests are given frequently and as a more or less regular part of the procedure during the class period as few as twenty or twenty-five items may be enough. This means, of course, that with such frequent use of short tests only a small amount of subject matter can be covered in a single test. If these examinations are given less frequently as, for example, every two weeks or once a month many more items should be included. Under these conditions one hundred items is probably a minimum. The importance of sampling extensively into the subject matter to be tested is demonstrated clearly by the diagram on page 77 in Chapter IV. A true and accurate measure of a pupil's information cannot be obtained by brief

[9] The term *item* is used to refer to the smallest unit or part of a test which calls for a distinct answer or pupil response. Thus a single direct recall or yes-no question, a single pair of expressions in a matching test, a single word to be defined or a single term to be connected with the proper part of a figure is an item. The sample above is one item in a true-false test.

tests of a limited number of exercises. This unreliability results from two factors: (1) the limited range of information sampled by the exercises; and (2) the chance which the pupil has of guessing the right answers to certain exercises without really knowing anything about them. The operation of the factor of chance is, of course, dependent upon the type of exercise used. Recall type exercises afford little opportunity for guessing. True-False, Yes-No, and all other types of alternate response exercises offer considerable opportunity for guessing. In multiple response exercises the chances for a pupil's guessing the correct answer are reduced as the number and apparent plausibility of alternate answers he has to choose from are increased.

52. Selection of the Subject-matter Items to be Tested. *Care should be exercised in the selection of the items to compose an objective test.* In the case of a limited field of subject matter in which the aims are quite clear and in which the items of information are specific practically all of the more important items can be included. In this case importance means that a complete understanding of the subject matter cannot be had without a knowledge of the items referred to. Where the field is broad the process of sampling, which has been previously discussed, must be resorted to in order to make sure that all aspects of the subject are covered in the examination.

53. Clear Statement of Exercises. *Statements should be carefully worded to avoid ambiguity.* The complete elimination of all ambiguity from objective exercises is extremely difficult to achieve. The following exercises illustrate this point:

True-False Exercises

A. Norms and Standards are often used synonymously. True False

B. The terms Norms and Standards are often used synonymously. True False

Exercise *A* creates some doubt in the mind of the reader as to whether it is the terms which are used synonymously or interchangeably, or whether it is the typical scores of pupils (norms) or the desirable goals of achievement (standards) which are used interchangeably. In the second exercise (Exercise *B*) this uncertainty has been removed by the addition of two words. The latter exercise is obviously quite superior to the former.

Objective exercises of the multiple response type in which a number of possible answers are suggested for choice should be so formulated that the correct answer is not too evident. In other words, the incorrect answers should not be so far-fetched or absurd that they can be recognized as such even without any special knowledge of the subject-matter being tested.

54. Difficulty of the Exercises. *The single items and also the whole test or examination should be of such a degree of difficulty that a minimum of perfect scores on the one hand and zero scores, on the other, will be made.* Examinations so difficult that no one is able to answer the exercises correctly do not reveal anything, since the examiner can never be certain how much information of difficulty just less than that of the exercises given, the subject does know. The same general condition exists when exercises are so easy that no one fails to answer them correctly. In such a situation there is no way to determine how far above the level of the difficulty of the examination the ability of the class or individual may be. Thus it is a generally accepted principle in educational measurement that test exercises which yield zero and perfect scores are to be avoided.

55. Arrangement of the Exercises. *The items included in an objective examination should be arranged in such a way that there is no regular sequence of answers.* For example, the true and false, or positive and negative, statements in an alternate response type of test should be arranged in

random order. The arrangement of exercises in any specific system might become apparent to the student and thus the entire examination would be invalidated. In the following comprehension exercises the answers are arranged in no special system. The grouping and sequence of Yes and No answers is purely a matter of chance which may be likened to the results of tossing a perfectly balanced disc or coin.

Sentence Comprehension Exercises

May external appearances be deceiving?	Yes	No
Should a person weed out bad habits before they are too firmly established?	Yes	No
May careful observation be of considerable aid in decreasing mistakes in one's conclusions?	Yes	No
Do controversies always result in the satisfactory solution of a problem?	Yes	No
Is it always impossible to misapply talent?	Yes	No
Are all laws enforced with facility?	Yes	No
May there be contention among the members of a parliament?	Yes	No
Are agility and endurance considered good qualifications for an athlete?	Yes	No
Are all festivities characterized by formality and ceremony?	Yes	No
Must exercise be violent to be considered adequate?	Yes	No
Are authorities sometimes quoted in editorials?	Yes	No
Are future events definitely predictable?	Yes	No

56. Arrangement of Answers for Convenience in Scoring. *The directions should require the pupils' responses to be recorded so that scoring is simplified.* For example, the practice of underlining the proper answer, as illustrated, makes it very easy to score the examinations rapidly and without error. The practice of having the pupils record the number of the statement or answer in the case of multiple choice exercises also simplifies the scoring process.

57. Use of Various Types of Exercises. *The principle of variety should be observed, that is, different types of exercises should be used from time to time in order to avoid the sheer monotony of always forcing the pupil to respond to the same type of exercise.* Furthermore, in a single long test it is usually better to use several types of exercises. For example, in an objective examination of 100 items in the field of history it might be well to have some such arrangement of exercises as this:

20 Recall exercises—Illustrated in Section 60, p. 106
20 True-False exercises—Illustrated in Section 60, p. 109
20 Multiple Choice exercises—Illustrated in Section 60, p. 107
40 Matching exercises
  Cause-effect (10 items)
  Date-event (10 items)
  Men-dates (10 items)
  Men-events (10 items)

Naturally the recall, true-false, multiple choice, and matching exercises will not be grouped together but will be distributed in accordance with some prearranged system throughout the test.

IV. Practical Use of Results

58. Use of Objective Test Scores as Basis for Class Grades. Pupils' scores from valid and reliable objective examinations afford the teacher's best single basis for measuring and grading pupil achievement within a given subject. The results of objective examinations take out the weakest features of the teacher's marking system. The validity of the objective examination depends upon the success of the teacher in formulating objective exercises which really represent the fundamental facts which the pupils have had an opportunity to learn. The reliability of the test is mainly a function of its length. This can easily be controlled. Teachers with a little practice can formulate exercises which are satisfactorily

objective. Thus it is apparent that practically the only point where the teacher's judgment needs to enter is in determining the approximate or relative difficulty of the examination as a whole so that the proportionate number of superior and inferior grades may be determined.

Objective test scores may be turned into the particular type of teacher's mark in current use by assigning high grades to a certain proportion of the higher scores, and low or failing grades to a certain proportion of the lower scores. The exact proportion of the class assigned certain grades is not always the same. Sometimes a very difficult examination may demand a slight shift in the teacher's standard of difficulty. The accompanying table (Table 3) shows the approximate percentage of pupils who in the long run should be assigned the indicated grades, assuming that the class as well as the examination is somewhere near typical.

Table 3

THEORETICAL BASIS FOR DISTRIBUTION OF OBJECTIVE TEST GRADES

	GRADE				
	A	B	C	D	Fd.
Approximate Percentages of Class.....	5	20	50	20	5

The assignment of class grades from objective test scores on this basis means that in a class of one hundred pupils approximately five with the highest scores will be assigned A's, the group of twenty with the next best scores will be assigned B's, etc. It must not be assumed that exactly these percentages of pupils will be assigned these grades within any given small class. A noticeably superior class may of course be given a higher percentage of high grades. Similarly a class which in the judgment of the teacher is below the

average may be given a smaller percentage of high grades. It is at this point that the judgment of the teacher must be used when his grades are based upon the results of objective examinations. The teacher is undoubtedly better able to judge the general level of class achievement than he is to apply his judgment reliably to each individual pupil as he is obliged to do with the traditional examination.

Class grades given in this manner for each such examination of the achievement of each pupil can be compared even though the examinations themselves are widely different in difficulty. The process of changing the examination scores into class grades takes care of this factor. The final grade which the pupil should be given for the subject, of course, can be readily obtained by taking the average of the grades assigned. Furthermore, if desired, there is the possibility of assigning a point value to each grade letter, computing the actual average grade given each pupil, and then re-assigning final grades on the basis of these averages. This is, in fact, a very convenient and satisfactory way of averaging and assigning final grades in large classes. When such a procedure is followed the point values given in Table 4 may be assigned

Table 4

SUGGESTED POINT VALUES CORRESPONDING TO CLASS GRADES

Grade	Points
A	10
B	7
C	5
D	3
Fd.	0

for the various grade letters. This table means, for example, that for each "A" a pupil is given 10 points, for each "B" he is given 7 points, for each "C" he is given 5 points, etc.

Assume that a pupil has the following grades on his record for a given course: One A, three B's, two C's, three D's, and one Failed. This gives him a total of 50 points for his ten grades, or an average of 5 points per grade. This is equivalent to an average of "C" if the same method is used in assigning the average grades.

V. Classification and Illustrations of Types of Objective Exercises

The more common types of objective test exercises may be classified into two main groups; (1) the recall type and (2) the recognition type. In the following pages a classification of the special types of test exercises falling into these two groups is given. Each of the more commonly used types of exercises is later illustrated by one or more exercises complete with directions for use. The varieties of test exercises named and illustrated do not constitute an absolutely complete list, but do include practically all of the types suited for general use. Many slightly different forms can be made by combining or mixing the features of two or more of the kinds given. Several of the types given illustrate such combinations or mixtures, but many more can be devised by teachers who will study these types of objective exercises.

59. Classification of Objective Tests. The main types of objective examinations which have been formulated and widely used to date fall into the following classifications:

I. Recall types
   1. Recall exercises or questions with single correct answers
   2. Completion exercises
      (*a*) simple completion exercises
      (*b*) completion exercises with suggested or controlled answers

II. Recognition types

1. Multiple response exercises
   (*a*) Multiple-Answer exercises with only one correct answer
   (*b*) Multiple-Answer exercises with answers of varying degrees of merit
   (*c*) Multiple-Answer exercises with one or more correct answers
   (*d*) Compound Multiple-Answer exercises
   (*e*) Multiple-description exercises
2. Multiple-Reason exercises
   (*a*) Multiple-Reason exercises with only one correct answer
   (*b*) Multiple-Reason exercises with reasons of varying degrees of merit
   (*c*) Multiple-Reason exercises with one or more correct reasons
3. True-false exercises
   (*a*) Single and Double Yes-No questions, answers to be underlined
   (*b*) Single and Double Yes-No questions, answers to be written
   (*c*) True-False statements, answers to be underlined
   (*d*) True-False statements, answers to be written
4. Incorrect statements
   (*a*) Alternative statements to be corrected
   (*b*) Incorrect statements
5. Matching exercises
6. Definitions
   (*a*) Simple definition or explanation
   (*b*) Same or opposites
   (*c*) Distinguishing exercises
7. Enumeration
   (*a*) Complete enumeration
   (*b*) Partial enumeration
8. Association exercises
   (*a*) Opposites
   (*b*) Genus-species
   (*c*) Connected terms
   (*d*) Disconnected terms
9. Exercises for arrangement in chronological order
10. Identification exercises

11. Analogies
    (*a*) With no answer suggested
    (*b*) With suggested answers
12. Abbreviations or formulae
    (*a*) Abbreviations, symbols, or formulae to be given
    (*b*) Abbreviations, symbols, or formulae to be expanded
13. Classification
    (*a*) Exercises with one extraneous word
    (*b*) Exercises with one or more extraneous words

60. Samples of Types of Objective Test Exercises. The examples given in the following pages have been chosen to demonstrate the manner in which the various types of exercises may be used in different school subjects. In fact, almost any of the kinds illustrated may be used in any one of the commonly taught school subjects. An analysis, however, makes it apparent that some are much better suited for use in certain subjects than in others. It is advisable in the use of all types of exercises described in the following pages and necessary in the use of most of them that mimeographed or otherwise printed copies of the tests be placed in the hands of the pupils. Even with mature students it has been found that exercises, such as the true-false type, can not be dictated to the pupils without reducing the reliability of the test.[10]

In selecting the samples no effort was made to choose the best possible ones in the various subjects. They are not based on any study of the subject matter commonly taught or the questions commonly asked. In other words, the examples in any one subject, if gathered together, do not constitute a test on the minimum essentials of that subject. So many textbooks and standard tests were consulted in an effort to secure suitable material for these illustrations that specific references to each source are not given.

[10] Paterson, D. G., "Constructing New Type Examinations," World Book Company.

A. Recall Types

(1) Simple Recall Exercises

This type, which is one of the easiest to prepare and which may be employed in practically all subjects, consists of questions having a single word or expression as the only correct answer. It may, therefore, be made perfectly objective if carefully prepared. The score is the number of correct responses.

Directions: Answer each of the following questions with a single word or expression. Write the word on the line immediately after the last word of the question.

1. Who was the President of the United States during the World War? _____
2. What is the product of x^2, x^3, and x^5? _____
3. What is the dative plural of homo? _____

(2) Completion Exercises

The completion type of exercise is one of the most commonly used of the newer objective forms. From the standpoint of pupil response it is probably one of the most valuable tests of this kind. It is very difficult, however, to construct a simple completion test so that the scoring is highly objective. Care should be taken to construct statements which require answers that may be definitely evaluated as either right or wrong. Completion tests are scored by counting the number of statements correctly completed or the number of blanks correctly filled in.

a. Simple Completion Exercises

Directions: The following statements are to be completed by writing one and only one word in each blank.

1. The ———— is a mammal.
2. Water boils at ——— degrees Centigrade.
3. Boyle's Law states the relation between the ———— of a gas and its ————.

b. Completion Exercises with Answers Suggested

Directions: Complete the following statements by inserting one of the words found in the list written on the right of the page. None of the words is to be written more than once.

1. ————	"Snowbound" was written by	1. Bunyan
2. ————	"Pilgrim's Progress" was written by	2. Coleridge
3. ————	"The House of Seven Gables" was written by	3. Whittier
4. ————	"The Ancient Mariner" was written by	4. Hawthorne

B. Recognition Types

(1) Multiple Response Exercises

Exercises of the multiple choice type are among the most satisfactory for use in the construction of objective examinations. With care in constructing and organizing the exercises the scoring can be made very objective. A very satisfactory practice is to number the proposed choices submitted for each exercise. The number of the proper choice can then be written on a line at the right margin of the test page which very decidedly simplifies the problem of scoring such tests. This is believed to be a distinct improvement over the practice of underlining the correct statement. This type is illustrated in the first sample given below.

a. Exercises with only one Correct Answer

Directions: Each of the following statements can be correctly completed by one and only one of the numbered expressions.

You are to write the number which stands for the correct expression on the line at the right of the exercise.

1. New York was settled by the (1) Spanish, (2) English, (3) Dutch, (4) French, (5) Portuguese. _____
2. The term of office of a United States Senator is (1) 1, (2) 3, (3) 5, (4) 6, (5) 7 or more years _____
3. The best temperature for living rooms is about (1) 60°, (2) 68°, (3) 75°, (4) 88°, (5) 78°. _____

b. Multiple-Answer Exercises with Answers of Varying Degrees of Merit

Directions: Underline the one word in the parentheses of each statement which makes the best statement.

1. (Copper, corn, paper, rubber) is a good conductor of electricity.
2. Maple is used for (casings, doors, floors, shingles).
3. Mahogany is used for (barrels, boots, boxes, crates, furniture).
4. The (pine, oak, elm, pear, ash) is a desirable shade tree.

c. Multiple-Answer Exercises with one or more Correct Answers

Directions: Underline all the words in each parentheses which will make true statements.

1. (Poe, Longfellow, Scott, Macaulay, Irving) were Americans.
2. (Webster, Grant, Patrick Henry, Jefferson Davis) were American statesmen.

d. Compound Multiple-Answer Exercises

Directions: Underline the one expression in each line which is closely related to the expression given in the first line.

1. Woodrow Wilson
   a. Massachusetts, New Jersey, New York, Pennsylvania.
   b. Vice-President, Senator, Ambassador, President.
   c. Republican, Democrat, Prohibitionist, Socialist.
2. Des Moines
   a. Illinois, Indiana, Iowa, Ohio.
   b. (Population) 100,000; 75,000; 15,000; 25,000.

e. Multiple-Description Exercises

Directions: Place a check mark (√) before the paragraph which best describes the city of Chicago.

1. Chicago, the largest city of Illinois, is one of the leading livestock markets of this country. It is also noted as a great railroad center. It is situated on Lake Michigan. It is often called the Windy City.
2. Chicago is a city of 3,000,000 population, located in the north central part of Illinois on the Chicago River. Situated in the center of a prosperous agricultural region, it is a leading center in the manufacture of flour and corn products.
3. The city of Chicago, with a population of more than a million, is located on Lake Michigan. Because of its location on the lake and because of its proximity to the mines of the northern states and its transportation facilities, it has become one of the centers of the corn and steel industry.

(2) Multiple-Reason Exercises

The characteristics of the multiple-reasons exercise are practically the same as those of the multiple-response type, and on this account samples are not presented.

(3) True-False Exercises

Because of the ease with which they may be administered and scored true-false exercises are very popular types of objective test exercises. The two types, those answered by either checking, underlining, or writing the words *true* or *false* after the exercise and those answered by using Yes or No

are identical in function although the exercises may be worded somewhat differently. Two main types of true-false or Yes-No exercises are recognized clearly. First, single statements, each covering a distinct item represent the most commonly used form. Second, double true-false statements, one true and one false covering each distinct test item. The first is the more commonly used form. The second is a newer type which has appeared as a result of an effort to devise a test which would automatically correct for the factor of guessing. The evidence shows that when this type of exercise is scored in terms of the number of pairs right the factor of chance is quite largely eliminated. Naturally in making an examination composed of these double true-false exercises the two exercises covering the same item are not allowed to appear close together. Experience shows that it is best to place the paired exercises in opposite sections of the test.

Since true-false exercises are fundamentally the same whether they be of the single or double type, the sample given as an illustration shows a series of paired exercises of the double Yes-No variety.

Directions to the Pupil: Study the samples below. Read each question and answer it by drawing a line under the right answer. DO NOT GUESS.

Samples: **Are all people dishonest?** YES NO

Are authors often quoted? YES NO

PART A

1 Should all students have the determination for achievement? ..... YES NO 1

2 Are all persons true to their best ideals? ..... YES NO 2

3 Do people sometimes co-operate for self preservation? ..... YES NO 3

4 May disputes ever arise over an expedient alliance? ..... YES NO 4

5 Does admiration always involve flattery? ..... YES NO 5

PART B

1	Do all students have the same determination to achieve?	YES	NO	1
2	Do many people fail to live up to their possibilities?	YES	NO	2
3	Does co-operation always result in self preservation?	YES	NO	3
4	Do expedient alliances always result in disputes?	YES	NO	4
5	May flattery be an expression of admiration?	YES	NO	5

The remaining types of objective exercises listed on pages 104 and 105 are so infrequently used in the construction of informal examinations of the objective type that they are not illustrated here in detail. Samples of practically all of the varieties of exercises not illustrated may be found in some form by examining any good collection of standardized test material.

61. Chance Factor in Objective Tests. By the very nature of their construction a great many objective test exercises involve the possibility of the pupil's guessing the correct answer by pure chance. In fact many teachers object to the use of the new type examinations because of this chance factor. This objection seems to imply that essay-type examinations do not possess the same weaknesses. When we consider that an essay-type examination may be assigned a grade varying from 20% to 90% depending on who graded it, there is certainly a large measure of chance present. Further, when we remember that an essay-type examination of ten or so questions always means a brief sampling from a large number of possible questions which the student may know or may not know, we see that the question is after all not one of using essay-type examinations and thus avoiding the operation of the chance factor. It is rather a question of using the type of exercises in which the operation of chance will be reduced as far as possible and at the same time permit

the use of the types of exercises which are better than the essay type for other reasons.

The chance factor is largely excluded in recall type exercises, but is present in practically all of the recognition types. The extent to which it is present depends upon the kind of exercise used. The chance factor can be more definitely controlled in the multiple-response type of informal tests than in the True-False or Yes-No types. In the case of the latter there is in theory at least a chance that the pupil may answer at least one-half of the exercises correctly by guessing.

The only question at this point is whether children really do guess under such circumstances or whether they are not almost invariably influenced in their responses by some residual knowledge of the answer and thus do not make a purely random guess. This situation has brought about the use of two types of instructions in the administration of these objective tests in the classroom. One group instructs the child to guess when uncertain, the other definitely instructs the child not to guess but to omit the answer to the exercise when uncertain. Practically all investigators in this field conclude that it is better to increase the length of the examination than to attempt to compensate for length by correcting for chance. Ruch [11] and Toops [12] have shown that correcting for chance decreases the reliability of the examination. On the other hand Ruch [13] has shown that it is better to instruct the pupils not to guess and then to apply the correction.

The common method of correcting two response and true-false tests for chance is to deduct the number of exercises answered incorrectly from the number answered correctly. For example, a child in taking a true-false examination answers 62 exercises correctly and 14 incorrectly. His corrected

[11] Ruch, G. M., "Improvement of the Written Examination." Scott, Foresman and Company, 1924.

[12] Toops, H. A., "Trade Tests in Education." Teachers College Contributions to Education, No. 115, 1921.

[13] Ruch, G. M., "Objective Examination Methods in the Social Studies." Scott, Foresman and Company, 1926.

score is 62 less 14, or 48. Scores of less than zero are not ordinarily assigned in tests of this type. Thus a child answering correctly only 24 exercises on the above test and making mistakes on 40 would receive a corrected score of zero. The correction for chance can be generalized into a simple formula which takes care of all types of tests without regard to the number of possible choices involved. The corrected score is equal to the number of exercises answered correctly minus the number answered incorrectly divided by the number of possible choices less one. In symbols this formula is as follows:

$$C = R - \frac{X}{N-1}$$

where R is the number of exercises answered correctly, X is the number answered incorrectly, and N is the number of choices.

VI. Summary

This chapter treats in an introductory manner the specific ways in which standard educational tests and the new type of examination or informal objective tests are unlike, and in what ways they are similar. A few of the major advantages and disadvantages of the objective examination are pointed out. In addition a number of general principles and suggestions for use in constructing and administering the objective type of examination are given. Among other practical suggestions a simple plan for the changing of objective test scores into class grades or marks is proposed and illustrated. Finally, in order to give the student and teacher a concrete notion of material of this kind representative samples of the more commonly used types of objective test exercises are given.

The brevity of this chapter makes it impossible for a great many important features of informal objective tests to be given adequate treatment. The teacher or student who wishes

to investigate this type of educational measurement more critically should study the references at the end of this chapter.

This and the preceding chapter, Chapter IV, have as their purpose the discussion of the practical problems involved in using standardized and informal objective tests in the classroom. Chapter VI discusses the problems involved in the summarization of test results for purposes of interpretation.

SELECTED REFERENCES

Asker, W., "The Reliability of Tests Requiring Alternative Responses." *Journal of Educational Research,* 9: 234–40, March, 1924.

Barthelmess, H. M., "Reply to a Criticism of Tests Requiring Alternative Responses." *Journal of Educational Research,* 6: 357–59, 1922.

Brinkley, Sterling G., "Values of New Types of Examinations in the High School." Teachers College Contributions to Education, No. 161. Columbia University, New York, 1924.

Chapman, J. Crosby, "Individual Injustice and Guessing in the True-False Examination." *Journal of Applied Psychology,* 6: 342–48, 1922.

Farwell, H. W., "The New Type Examinations in Physics." *School and Society,* 19: 315–22, March, 1924.

Greene, Charles E., "New Type Tests." Research Monograph No. 3, Denver, Colorado, 1926.

Greene, H. A., "A New Correction for Chance in Examinations of the Alternate-Response Type." *Journal of Educational Research,* 17: 102–107, February, 1928.

Hahn, H. H., "A Criticism of Tests Requiring Alternative Responses." *Journal of Educational Research,* 6: 235–40, 1922.

Knight, F. B., "Data on the True-False Test as a Device for College Examinations." *Journal of Educational Psychology,* 14: 123–124, 1919.

Monroe, W. S., "Written Examinations and Their Improvement." Bulletin No. 20, University of Illinois, 1922.

Paterson and Langlie, "Empirical Data on the Scoring of True-False Tests." *Journal of Applied Psychology,* 9: 339–48, 1925.

Paterson, Donald G., "Preparation and Use of New Type Examinations." World Book Company, Yonkers-on-Hudson, New York, 1925, pp. 70–71.

Remmers, H. H., "An Experimental Study of the Relative Difficulty of True-False, Multiple Choice, and Incomplete Sentence Type of Examination Questions." *Journal of Educational Psychology,* 14: 366–72, 1923.

Ruch, G. M., "Improvement of the Written Examination." Scott, Foresman and Company, Chicago, 1924. Chapters IV and V.

Ruch, G. M., and Stoddard, G. D., "Comparative Reliabilities of Five Types of Objective Examinations." *Journal of Educational Psychology,* 16: 89–103, 1925.

Ruch, G. M., et al., "An Investigation of Objective Examinations in the Social Studies." Scott, Foresman and Company, Chicago, 1926.

Ruch, G. M., and Stoddard, G. D., "Tests and Measurements in High School Instruction." World Book Company, Yonkers-on-Hudson, New York, 1927.

Russell, Charles, "Classroom Tests." Ginn and Company, Boston, 1926.

West, P. V., "Significance of Weighted Scores." *Journal of Educational Psychology,* 15: 320–26, 1924.

West, P. V., "A Critical Study of the Right Minus Wrong Method." *Journal of Educational Research,* 7: 1–9, 1923.

CHAPTER VI

SUMMARIZING THE RESULTS OF TESTING

This chapter has as its purpose the discussion of the following points to be considered in the interpretation of test results:

a. The most important statistical procedures which teachers need to use in summarizing test results for interpretation.
b. The tabulation of test scores from large numbers of pupils for convenience in summarizing the results.
c. The common measures which may be used to express the typical or average achievement of a class, and the methods by which these measures may be computed.
d. The common measures by which the spread or variation of test scores may be expressed; why they are needed and how they are computed.
e. The methods of using test scores as a basis for ranking or placing pupils in order of ability or merit.

62. Introduction. General observation of the work of pupils, and experience with the use of tests in the classroom indicate that a wide range of accomplishment may be expected from the different individuals in a given class. Some methods of summarizing these widely varying results, so that they may accurately and simply be described, are necessary. Summaries and descriptions of this type are statistical problems, but this fact need not disturb the student or teacher, for after all most of these elementary statistical procedures are simple. The main difficulties are the learning of a new and different type of vocabulary, and the revival of a few relatively simple arithmetical skills.

The most important statistical techniques from the standpoint of the frequency of their use in the classroom are: (1) a knowledge of how to classify and tabulate data; (2) a knowledge of how to compute the common measures of central tendency; (3) a knowledge of how to express the variability or spread of the data; (4) a knowledge of how to express the relationship between two or more groups or arrays of data; (5) a knowledge of how to secure and use standards, norms, and derived scores for purposes of comparison and interpretation of test results.

The discussion of these items constitutes the major portion of this and the following chapter. This relatively large amount of emphasis is given to these points for two reasons: (1) successful and satisfactory work with tests can be expected only when the person using them is adequately equipped to understand and interpret them. Such abilities are dependent upon a reasonable mastery of these elementary statistical techniques. (2) Current educational literature in practically all fields is literally filled with the terms and the techniques discussed here. Reports of progress in education are dependent upon statistical methods. If the teacher and the student are to keep abreast of the time educationally, the ability to read statistical discussions in current educational literature with understanding must be developed.

I. Classification and Tabulation of Test Scores

63. Need for a Method of Grouping Data. The very fact that people are unlike physically and mentally gives rise to the need for statistical methods. This is illustrated by the history test scores given in Table 5. Pupil A, an eighth grade pupil, made a score of 52 points on a certain history test. Pupil J in the same school grade made a score of only 20 points. An examination of the data in this table shows that other pupils made scores between these two extremes. If this were a large class and the range of scores (the difference of the highest and the lowest scores) were greater it would

be very difficult for a teacher to secure a clear picture of the actual achievement of this class as represented by these scores. As a matter of fact the arranging of scores in order of size is one of the very simplest steps in summarizing test results for interpretation. It enables the teacher to pick the highest and the lowest scores easily, and also after some practice to pick from the scores the single one that seems to be most typical of the group. Clearly, such a typical score cannot be one of the extreme scores, therefore one is almost forced to look toward the middle of the group for such a representative score. This point is discussed in detail in Section II,[1] page 127 of this chapter.

Table 5

SCORES ON A HISTORY TEST TAKEN BY AN 8TH GRADE CLASS

Pupil	Score
A	52
B	50
C	47
D	41
E	36
F	36
G	30
H	27
I	21
J	20

The relatively simple practice of arranging test scores in order of size from highest to lowest or the reverse is helpful, however, only when the number of pupils in the class is not large. When more than ten or so cases are present it becomes more helpful, perhaps even necessary, to arrange the scores into groups more compactly so that they may be easily read and comprehended. This process of grouping the cases into

[1] See also Greene, H. A., "Work-Book in Educational Measurements." Work-Unit Number I, Problems 1 to 4 inclusive.

the form of a table is called *making a frequency distribution* and the resulting table is called a *frequency table.* The need for this method of grouping scores is apparent after examining the scores of the thirty-seven 7th grade pupils on an arithmetic test given in Table 6. It is impossible to grasp the significance of such a large number of widely scattered test scores without resorting to grouping.

Table 6

ARITHMETIC SCORES OF THIRTY-SEVEN 7TH GRADE PUPILS

(As Determined by an Arithmetic Test)

72	71	68	68	66	66	50	60	48	46
46	45	43	25	52	53	53	54	55	55
63	58	57	40	38	40	36	34	30	32
30	28	24	70	60	64	40			

The method of grouping test scores is similar to that followed by postal clerks in the distribution of letters and other mail in a large railway postal terminal. Mail designated for certain sections of the country or for certain large centers from which it is re-distributed is thrown into the proper mail pouches. Each pouch may, however, contain mail addressed to a number of postoffices in the same section of the country. For example, in the Chicago postal terminal mail addressed to postoffices in the Pacific north-west may be consigned to a certain group of pouches, while mail going to the south-east section of the United States will be consigned to other pouches representing that section of the country. The number of pouches required depends upon the number of pieces of mail to be distributed and also upon the size of the section of the country which can be most efficiently served by a given pouch. Increasing the number of pouches naturally increases the labor involved in sorting the pieces of mail, but at the same time it increases the accuracy of the distribution. Mail in Chicago might be sorted into two classes—eastbound and westbound. This would in-

troduce a large error since not all sections of the country would be effectively served by this classification.

64. Steps Involved in Preparing a Frequency Table. The purpose of the foregoing extensive illustration is to give the reader a clear conception of the purposes and the problems involved in grouping test scores into a frequency table. In the data given in Table 6, page 119, the scores range from the low score of 24 to the high score of 72. The process of grouping data involves the following three steps:

(*1*) The determination of the range of the scores. This is similar to determining how far apart the sections of the country are which are to be served by the postal terminal station. This aids in determining how many mail pouches must be used.

(*2*) The determination of the number and the size of the groups in the classification. If a given mail pouch is to serve a great many minor postal stations a larger type of pouch and a smaller number of pouches may be used than if each minor postoffice has its own individual pouch.

(*3*) The tabulation of the scores. The pieces of mail can not be sorted accurately and effectively until the number of mail pouches and the exact sections of the country served by each have been determined. After this has been done the trained postal clerk can accurately and rapidly sort the pieces of mail into their proper pouches.

The result of applying these three specific steps to the test scores given in Table 6, page 119, is given in Table 7, on page 125. This is representative of the type of frequency table around which the following discussion will center.

65. Determining the Range of Scores. The *range* of scores is found by subtracting the smallest score from the largest score. In the case of the scores given in Table 6, page 119, the range is the difference of 24, the lowest score, and 72, the highest score. The range is therefore 48. The range is useful in determining the number of class-intervals or steps to be used in making the frequency table.

66. Determining the Number and Size of Class-Intervals. Since the scale of a frequency table must include the entire range of scores present the size of the class-interval used depends upon the range of the scores, and upon the number of such class-intervals, or steps, which it seems wise to use in the table. No special rule relative to the number of steps or class-intervals to be used in a frequency table can be stated. However, it is probably safe to state that it is unwise to group data into fewer than twelve or fifteen class-intervals because of the *error of grouping* as the number of steps is decreased. This point was suggested in connection with the illustration of the sorting of mail in the terminal station. If all mail westbound from Chicago were placed in three pouches consigned to St. Paul, St. Louis, and Omaha, considerable injustice would be done to other sections of the country. The more the number of pouches is increased the more accurate the resulting routing of the mails—up to the point where so many separate pouches are used that they can not be handled economically. Similarly, there are a number and a size of class-interval which seem to be best for each distribution.

The error of grouping of scores is naturally greater where only a few cases are found. It is also greater when the scores progress irregularly along the scale. For example, the tabulation of scores such as 38, 38, 38, 38, 40 into a three-unit class-interval of which 39 is assumed to be the mid-point introduces an error of grouping. In placing these five scores in the same class-interval it is assumed that they will together represent an average value which is the same as the mid-point of the step. The error in this case is not particularly serious. The average of the five scores is 38.4, instead of being 39. In tables with larger class-intervals it must be obvious that the error may be much greater.

It is probably unnecessary to use more than eighteen to twenty steps or class-intervals in making a frequency table in practical work with educational test data. The use of more than twenty steps involves somewhat more labor in making the distribution, and experience shows that where

there are enough cases to warrant making a frequency table the use of more class-intervals does not greatly affect the error of grouping involved. The main idea of grouping the scores into approximately 15 steps as suggested is to organize the scores into sufficiently small groups that they may be thought about effectively and yet not classify them into so few groups that important differences are covered up or significant errors are introduced.

When the two factors (1) the range of the scores, and (2) the approximate number of class-intervals which is to comprise the table are known, the size of the class-interval itself can be determined. The range of the scores found by taking the difference between the largest and the smallest scores in the array, gives a suggestion as to the number of units which may be used in each step or class-interval, assuming that the table is to have approximately 15 steps. Let us refer again to the scores given in Table 6, page 119. The range is from 72 to 24 points, or 48 points. If the range is divided by the number of class-intervals desired the resulting quotient indicates that the size of the intervals in this case should be slightly more than three units if only fifteen steps are used. However, the number of steps is somewhat flexible, so that for the data used in this illustration the use of class-intervals of three units each is desirable. (See Table 7, page 125.) Experience with the tabulation of test scores indicates that class-intervals of one, three, five, etc., units are more convenient for use than steps of two, four, six, etc., units. Intervals of one, three, five, etc., units simplify later calculations by making the mid-points of the steps conveniently whole numbers. By using intervals with two, four, and six units the mid-points are not only not whole numbers, but the limits of the class-intervals themselves are also fractional. This point will be discussed further in a later section in this chapter.[2]

67. Determining the Limits of the Class-Intervals. The next

[2] *Ibid.*

step in setting up the class-intervals for a frequency table is that of establishing the *limits* of each interval. In order to make this more easily understood a few terms must be defined. The *class-interval* or *step* is the group, or compartment within the limits of which given scores are assigned. The *mid-point* of the step or interval is a point mid-way between the upper and the lower limits of the step. For convenience in tabulation it is found desirable to choose the limits of the step in such a way that this mid-point is a whole number.[3] This, of course, makes it necessary that the upper and lower limits of the step be fractional values whenever odd-sized steps are used, but this is thought by many statisticians to be desirable also for the following reason: In most educational measurement there is the possibility of fractional scores, or of work on exercise units partially completed. It is thought that a score of 42 may as often really represent a score of 41.5 as it does a score of 42.5. The most likely score representing scores varying from 41.5 to 42.5 is of course 42. This method has the additional merit of giving a natural location on the scale to all scores expressed in whole numbers. A further advantage is that it makes it possible to compute a median (explained and discussed later in this chapter) from grouped data which is practically identical with the median found from ungrouped data.

It is now possible to proceed to set up the class intervals for the frequency table. In the case of the data given in Table 6, page 119, it is clear that the table must extend over a range of 48 points. To do this the table must provide

[3] At this point attention should be called to a rule of tabulation which if followed by members of large classes will greatly simplify the work of checking the papers. The rule is as follows: Select as mid-points for the class-intervals values which are divisible by the number of units per step. For example, in frequency tables using three unit steps the mid-points should be multiples of three, as 18, 24, 27, 30, etc. If every student follows this rule the tabulations will all be made in the same manner and may be checked from one answer key. It is from this standpoint alone that this suggestion is made and not because of any statistical merit which it may have. The answers to the problems in the *Appendix* of this book are worked on this basis.

class-intervals or steps high enough on the scale to accommodate scores of 71 and 72. At the same time steps must go low enough on the scale to accommodate scores as low as 24. The procedure at this point is, therefore, to begin at one end or the other of the distribution and build up the steps systematically. For example, if we start at the bottom with the lowest score in Table 6 it will be necessary to select as the mid-point of the step the value 24 in order that it may go low enough on the scale and at the same time be divisible by the number of units in each step. Since the interval is comprised of three units with 24 as the mid-point all scores of 23, 24, and 25 will go in this interval. In order, however, that all possible values of 23, 24, and 25 may be accommodated within the interval the limits of the interval must be established as 22.5–25.5. The accompanying diagram (Figure 3) shows this.

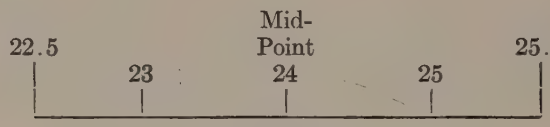


FIGURE 3

The results of applying these principles to the construction of a frequency table are shown in Table 7 on page 125 which is the frequency distribution of the Arithmetic Scores given in Table 6 on page 119.

68. Tabulating the Scores. The tabulation of scores is nothing more than a filing or pigeon-holing process which places within the same class-intervals scores which fall within the limits of these intervals. Table 7 shows that there were two scores which fell within the limits of the interval of 70.5 to 73.5; that there were 3 scores which fell in the interval 67.5 to 70.5, etc. The appearance of a score in an interval is usually indicated by a short vertical mark. The presence of five scores in a step is indicated by four vertical marks (or check marks) crossed off by a fifth mark drawn diag-

onally through the four, as in the case of the interval where the mid-point is 54 in Table 7. The number of check marks

Table 7

SCORES GIVEN IN TABLE 6 GROUPED IN INTERVALS TO FORM A FREQUENCY DISTRIBUTION

(1) Class Intervals	(2) Mid-Points	(3) Tabulated Scores	(4) Frequency (*f*)	(5) Deviation (*d*)	(6) *fd*	
70.5–73.5	72	//	2	+7	+14	
67.5–70.5	69	///	3	+6	+18	
64.5–67.5	66	//	2	+5	+10	
61.5–64.5	63	//	2	+4	+ 8	
58.5–61.5	60	//	2	+3	+ 6	
55.5–58.5	57	//	2	+2	+ 4	
52.5–55.5	54	~~////~~	5	+1	+ 5	+65
49.5–52.5	51	//	2	0		
46.5–49.5	48	/	1	−1	− 1	
43.5–46.5	45	///	3	−2	− 6	
40.5–43.5	42	/	1	−3	− 3	
37.5–40.5	39	////	4	−4	−16	
34.5–37.5	36	/	1	−5	− 5	
31.5–34.5	33	//	2	−6	−12	
28.5–31.5	30	//	2	−7	−14	
25.5–28.5	27	/	1	−8	− 8	
22.5–25.5	24	//	2	−9	−18	−83
Total Number of Cases (*N*).....	37		37			

Total Frequencies	37
Mean Ungrouped Data	49.78
Mean Grouped Data	49.53
Median	51.75
Range	48
Upper Quartile	61.13
Lower Quartile	38.44

in each step is then usually totaled and entered under the column marked "frequency."

69. Summary of Steps in Making Frequency Table. The classroom teacher will very often find the construction of a frequency table unnecessary in his work with tests since he usually works with small numbers of cases and can check the scores from the papers themselves. However, there are many occasions when the frequency distribution is necessary. It affords a very effective way of recording and preserving the results of using tests in the classroom. Furthermore, it affords the means for the use of a number of short cuts in the calculation of certain statistical measures useful in interpreting test results. The methods by which these measures are computed are given in succeeding sections of this chapter. This section summarizes in concise form the steps necessary in making a frequency table.

(1) Determine the range of the scores by subtracting the smallest score in the series from the largest score.
(2) Determine from the range of the data the number of class-intervals which should be used. No special rule covering the number of class-intervals can be given. However, it is probably safe to state that it is usually unwise to group data into fewer than twelve steps and probably unnecessary to group data into more than twenty steps.
(3) Use a class-interval of one, three, five, etc., units in preference to one of two, four, six, etc., units. The following table may be a helpful guide in your first experience with this type of work.

Table 8

Relation of Range and Size of Class-interval

For a Range of	Use a Class-Interval of
10 to 20	1
21 to 60	3
61 to 100	5
101 to 150	7
151 to 300	15

(4) Select as mid-points of the class-intervals values which are multiples of the size of step used. If a step of three units is used select mid-points which are divisible by three, etc.
(5) Set up the mid-point values of the intervals required and establish their limits clearly, using the fractional values as limits.
(6) Tabulate the scores determining the frequency of scores in each class-interval and the total frequency of scores in the entire table, as illustrated in Table 7, page 125.[4]

II. Measures of Central Tendency

The grouping of test scores into frequency tables is one step in the process of condensing them so that they can be analyzed and interpreted. However, something further must be done before it is possible to describe the table itself. Some single term or value which is representative of the entire table must be found. Since these values which may be taken to represent an entire distribution of scores are usually found near the center of the data when arranged in order of size they are commonly called *measures of central tendency.* The three common measures of central tendency are: (1) the arithmetic mean; (2) the median; and (3) the mode. Of these three measures of central tendency the median and the arithmetic mean are the most commonly used in educational measurements, and are accordingly the only ones emphasized in this discussion.

70. Computing the Arithmetic Mean—Ungrouped Data. Practically everyone uses and knows how to find the so-called average or *arithmetic mean.* It is commonly defined as the

[4] Extensive practice on the tabulation of data is necessary before the student will find it easy. For this purpose the student is referred to Problems 1, 2, and 3, on pages 357–358 in the *Appendix* of this book, and to a "Work-Book in Educational Measurements" by H. A. Greene.

This *Work-Book* consists of a series of forty-five practical problems in the tabulation, analysis, and interpretation of test scores. See particularly Problems 1, 2, 3, and 4 of Work-Unit I.

measure resulting from dividing the sum of the measures in the distribution by the number of measures. In the case of the scores shown in Table 1 page 51 the arithmetic mean is found by taking the sum of the scores earned by all of the pupils involved and dividing this by the number of pupils in the class. The sum of the scores earned by pupils A, B, C, etc., is 360. There are ten pupils in the class, therefore 360 is to be divided by 10. The result, the arithmetic mean, is 36. The value of this measure lies in the fact that it lends itself to describing by means of a single term a group of widely varying scores or measures. It expresses in very compact form one specific fact about the scores in which each single score has a part. On this account it is one of the basic statistical measures of central tendency.

71. Computing the Arithmetic Mean—Data Grouped in Frequency Distribution. The arithmetic mean can also be readily found from a frequency distribution. In order to make the procedure somewhat more definite it is advisable to redefine the term *arithmetic mean.* When considered from this point of view *the arithmetic mean is defined as a point on the scale, such that the sum of the deviations of the values larger exactly equals the sum of the deviations of the values smaller than it is.* Expressed in physical terms it may be considered the point at which the fulcrum must be placed in order to balance the scale when it is considered as a beam of regularly increasing thickness or density. For example, a beam of a certain length, two inches in thickness and varying in width from two to seven inches, can be balanced by placing a fulcrum at the proper point. This point may be determined experimentally or by mathematical calculation. Without regard to the method employed the fulcrum must be so placed that the moments of the forces on one side are exactly equalled by the moments of the forces on the other. Those who have forgotten the meaning of the term *moments of force* will be reminded by the statement that a force of three pounds resting on a lever one foot in length is exactly equalled by

a force of one pound resting on a lever three feet in length. (See Figure 4.)

The parallel between the physical lever and the mathematical calculation of the arithmetic mean is quite close. The problem in each case is to balance the forces on either side of a point to be determined. In case the physical lever is out of balance the correction is made by moving the fulcrum in the direction of the heavier end until equilibrium is established. In calculating the arithmetic mean a sort of trial balance is taken. If the moments of force are too great on one side the point of rotation is similarly moved in the direc-

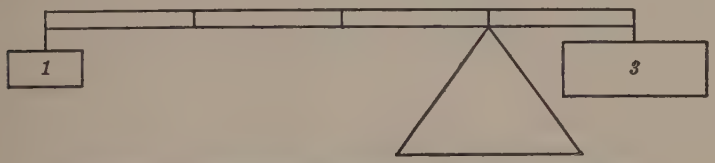


FIGURE 4. Illustrating Principle of Moments of Forces.

tion of the heavy end until the difference between these two forces becomes zero. This may be aptly illustrated by taking a procedure which classroom teachers have undoubtedly frequently used. It is necessary to obtain the average of five class marks for various periods through the school year in arriving at the final mark. These grades may be taken for illustration as 93, 90, 89, 88, 86. By inspection it may be seen that 89 is approximately the correct mean. The 90 is one point too large, the 93 is four points too large. In a corresponding way 88 is one point too small and 86 is three points too small. The total of the differences above the mean is five and the total of the differences below the assumed mean is four; therefore the assumed mean of 89 is too small by the amount of this difference (1) divided by the number of cases (5). This is equal to .2, and the mean is 89.2. This checks exactly with the mean found by the method of totaling the measures and then dividing by the number of measures.

This principle may be applied almost directly to the cal-

culation of the arithmetic mean from the frequency table. A sort of trial balance is taken by looking at the array and assuming as a mean some value on the scale which looks about right. As a matter of fact it is not necessary that this assumed point should be even approximately near the true mean. Any point on the scale could be assumed with equal accuracy. For the sake of illustration refer to Table 7 page 125 where the step marked "0" in the deviation column will indicate the point taken as the assumed mean. This point is called *zero* since all the measures in that particular step are assumed by this process to deviate zero from the mid-point of the step. All deviations on either side of this point start from here. Deviations in terms of steps which are representative of the intervals on the scale, comprising in this case three units, are laid off above and below. The steps in the ascending order of magnitude in the scale are designated as plus. Those in the descending order are designated as minus. The total magnitude of forces on the upper half of the scale is found by multiplying each frequency by its corresponding step deviation and taking the sum of these products (marked in Table 7 as *fd's*). The sum of the plus *fd's* in this calculation is 65. The sum of the minus *fd's* is 83. The algebraic sum of these two quantities is —18 (65 and —83 = —18), which expresses approximately the force tending to rotate the scale in the direction of the greater magnitudes. In order to bring about an exact balance of these two forces, the assumed mean which was selected at the outset must be moved slightly in the direction of the heavier end of the scale which is in this case in a minus direction. This is illustrated graphically in Figure 5. Since there are 37 measures in the distribution and each of the 37 measures contributes to the resultant force of 18 units, the average correction in this case is the result of dividing —18 by 37 which gives a correction of approximately —.49 steps. Since each step is based on three units it is necessary to multiply —.49 by 3 in order to turn it into scale units. The resulting value of —1.47 units represents the exact

amount by which the assumed mean must be corrected. The sign of this value resulting from the algebraic addition mentioned above is negative, therefore, the correction must take place in the negative direction.

When the assumed mean was taken the step represented by the values 50, 51, and 52 on the scale was tentatively accepted as that point. Since there are possibilities for a number of values to fall within this step in making the frequency table it becomes necessary for some one value to be assigned as most nearly representing the step. It is the practice in this procedure to take as this value the *mid-point* of the step. In this case the mid-point is 51. Since by the previous calculation it has been shown that the cor-

FIGURE 5. Illustrating the Calculation of the Mean.

rection for this assumed mean must be made in a negative direction the value of 1.47 must be subtracted from 51 which leaves 49.53 as the arithmetic mean of this distribution calculated by means of the shorter method.

The actual arithmetic mean of the 37 scores in this frequency distribution found by the long process is 49.78. The student will do well to recheck all of the work of finding this arithmetic mean from the data in Table 7 in order to become familiar with the various steps in the computations.

72. Summary of Steps in Finding Arithmetic Mean. The following is a summary of the specific steps involved in calculating the arithmetic mean from a frequency distribution by the shorter process.

1. Take the mid-point of some likely step on the scale as the *assumed mean.* Call this assumed point *zero.*
2. Lay off *steps of deviation* above and below this assumed mean maintaining the usual algebraic signs.
3. Find the *fd's* by multiplying the frequencies by the deviations. Carry the signs.
4. Find the algebraic sum of these *fd's* keeping the sign of the result.
5. Divide this resultant value by the number of cases in the distribution, and multiply this value by the number of units per step. This result is the *correction* (c).
6. In accordance with the sign of this correction (c) add to or subtract from the value assumed as the mid-point of the assumed mean of the scale (the step taken as zero). This should give the *arithmetic mean.*

73. Classroom Use of the Arithmetic Mean. The practical usefulness of the arithmetic mean in connection with the daily work of the classroom teacher is illustrated in the case of Miss L., a fifth grade teacher. At the end of a three weeks' period of instruction over a certain unit of geography Miss L. gave an objective informal test of the true-false type which had been given to similar classes in previous years. She wished to compare the achievement of this class with that of the classes which had taken the test previously. Some measure of central tendency was necessary. She therefore prepared a frequency distribution showing the scores earned by the 42 pupils in her class, and from this table computed the arithmetic mean of the scores. This arithmetic mean she was able to compare with the results of using this test in previous years. The frequency table and a copy of her work showing the computation of the arithmetic mean step by step is given in Table 9. The student will do well to check this work carefully and see that all calculations are correct.[5]

[5] A large amount of practice is needed in order to develop speedy and accurate work in computing arithmetic means. See Problems 4, 5, 6, and 7, on pages 361 to 363 of the *Appendix* for drill on this type of work. For a large amount of material specially designed for this purpose see Greene, H. A., "Work-Book in Educational Measurements," particularly Problems 5 to 10.

Table 9

DISTRIBUTION OF SCORES AND THE CALCULATION OF THE ARITHMETIC MEAN OF THESE SCORES OF A TRUE-FALSE TEST BY MISS L.'S FIFTH GRADE CLASS IN GEOGRAPHY

Class intervals	f	d	fd
112.5 (115) 117.5	1	+6	+6
107.5 (110) 112.5	0	+5	+0
102.5 (105) 107.5	1	+4	+4
97.5 (100) 102.5	3	+3	+9
92.5 (95) 97.5	6	+2	+12
87.5 (90) 92.5	4	+1	+4
82.5 (85) 87.5	9	0	(+35)
77.5 (80) 82.5	5	-1	-5
72.5 (75) 77.5	3	-2	-6
67.5 (70) 72.5	2	-3	-6
62.5 (65) 67.5	3	-4	-12
57.5 (60) 62.5	1	-5	-5
52.5 (55) 57.5	2	-6	-12
47.5 (50) 52.5	1	-7	-7
42.5 (45) 47.5	1	-8	-8
Total (N) = 42			(-61)

Step 1. Assumed Mean = 85

Step 2.

Step 3.

Step 4. +35 - 61 = -26

Step 5. $\frac{-26}{42} = -.619$

$-.619 \times 5 = -3.095$

Step 6. 85.000
3.095
81.905 or 81.91

My class average last year on this same test was 70.2 points, showing a slight superiority in the achievement of this year's class.

74. The Median as a Measure of Central Tendency. The arithmetic mean which has been discussed in foregoing sections of this chapter is one of the most popular and useful measures of central tendency for most research and experimental purposes. Its present popularity is undoubtedly due to the simplified methods of computing it which have been developed. However, in the interpretation of test results the median is still, and will probably always continue to be widely used.

Early workers with tests popularized the practice of taking the score of the middle paper of a pile of test papers arranged in order of size of scores as the expression of the central tendency of the group. It is the ease with which this so-called

median is found that has appealed to the classroom teacher. For a long time this was called the *median*. However, more recent workers with tests have recognized the fact that the score of the middle paper in a pile of test papers stacked in order of size of scores is not the same as the middle point on the scale of a frequency table of these same scores. Accordingly a distinction is now made between the score found on the middle paper of a pile of stacked papers and the median proper. The score of the middle paper of a pile of test papers arranged in systematic order is called the *mid-measure* to distinguish it from the *median*, which is the corresponding value when the data are grouped in a frequency distribution. Thus the *mid-measure* is a counting median found from ungrouped data. The *median* is computed from tabulated data.

75. Computing a Counting Median or Mid-Measure. The method of computing the mid-measure is illustrated by referring to the data given in Table 6, page 119. The arithmetic test scores as given in Table 6 are not arranged in any systematic order. The scores from these 37 test papers are arranged in descending order in Table 10.

Table 10

ARITHMETIC TEST SCORES GIVEN IN TABLE 6
ARRANGED IN DESCENDING ORDER

72,	71,	70,	68,	68,	66,	66,	64,	63,	60,	60,	58,	57,	55,
55,	54,	53,	53,	52,	50,	48,	46,	46,	45,	43,	40,	40,	40,
38,	36,	34,	32,	30,	30,	28,	25,	24.					

The mid-measure is the score of the middle paper, or the average of the two scores nearest the middle when the number of cases is even. In this case the number of papers or scores is 37 (odd). Thus the mid-measure is 52, a score such that

there are just as many equal to or larger as there are equal to or smaller than it is.[6]

76. Computing the Median from Frequency Distribution. By definition the mid-measure and the median are quite similar, the main distinction being that the *mid-measure* is designated as *an actual score on a certain paper* (or the average of the scores on two papers lying nearest the middle paper), while the *median* is defined directly in terms of *a point on the scale* of the frequency table on which it is based. *The median is a point on the scale such that 50% of the cases in the distribution are above it and 50% of the cases are below it.*

At this point it will be helpful if the student will recall that in working with the arithmetic mean from data grouped by class-intervals it was assumed that the scores falling in a given step were to be considered as having the value of the mid-point of the step. This amounts to assuming that all of the cases in a step are *concentrated at the mid-point* when computing the arithmetic mean. However, in computing the median a different assumption is made. It is assumed in this case that regardless of how few or how many cases fall in a given step these cases are *distributed uniformly through the step.* Thus, if there are three cases in a given step, one of the cases represents exactly one-third of the step. If there are 15 cases in the step one case represents one-fifteenth of the step. If this point is kept in mind much of the difficulty of finding the median will be removed.

The method of calculating the median is illustrated by the data given in Table 6, page 119, which is reproduced here as Table 11 in order to show all of the steps in the work. The following statements summarize the processes accurately as worked out in this illustration. To find the median of a series of scores arranged in frequency distribution:

[6] Additional practice in finding the mid-measure is given in Problems 8, 9, and 10 in the *Appendix* of this book and in Problem 11 of the "Work-Book in Educational Measurements," by H. A. Greene.

(1) Divide the total number of measures by two to determine 50% of the cases. In this table the total number of measures (N[7]) is 37. Thus, $N/2$ equals 18.5, or the half-sum.
(2) Beginning at the bottom of the column of frequencies count up as far as possible without exceeding the half-sum, 18.5. In this case the frequencies $2 + 1 + 2 + 2 + 1 + 4 + 1 + 3 + 1$ give a total of 17 without exceeding the half-sum.

Table 11

ARITHMETIC SCORES OF A SEVENTH GRADE CLASS
(From Table 6)

Class Intervals	f	
70.5–73.5	2	$\frac{N}{2}=18.5$
67.5–70.5	3	
64.5–67.5	2	
61.5–64.5	2	Sub-total = 17
58.5–61.5	2	
55.5–58.5	2	
52.5–55.5	5	$\frac{1.5}{2.0}\times 3 = 2.25$
49.5–52.5	2	
46.5–49.5	1	
43.5–46.5	3	
40.5–43.5	1	
37.5–40.5	4	49.5
34.5–37.5	1	+2.25
31.5–34.5	2	51.75 = Median
28.5–31.5	2	
25.5–28.5	1	
22.5–25.5	2	
	$N=37$	

[7] N means the total number of cases when referring to a frequency table.

(3) The difference between the *half-sum,* 18.5, and the *subtotal,* 17, shows how far into the class-interval in which the median is located it is necessary to go in order to come to the median itself. Subtracting 17 from 18.5 leaves 1.5 units which must be taken from the step in which the median is located. Since the next step contains two cases the median is located 1.5/2.0 or three-fourths of the way through the step.

(4) The two cases in the step are considered to be uniformly distributed through the interval; hence the determination of the correction, or the fraction of the step which must be passed in counting up the frequencies to a place where 50% of the cases are found below, is merely that of taking whatever portion 1.5/2.0 is of the number of units per step. There are three units in each step in this table, therefore the correction equals $1.5/2.0 \times 3$, or 2.25 units.

(5) Since the correction is to be made into the interval beginning 49.5 this value 2.25 units is to be added to 49.5. The median thus becomes 51.75. Obviously, if these calculations had been made by adding the frequencies down from the top of the table the same median would have resulted, but the correction would be .75 units and would be deducted from the top of the step instead of being added to the bottom of the interval in which the correction is made.[8] $(52.5 - .75 = 51.75)$. This fact in itself affords a very convenient check on the calculation of medians. The student will find that it is an excellent practice to check the accuracy of all calculations in these illustrations.

77. Use of the Median. The actual use of the median as a measure of central tendency for the comparison of test scores from a first grade reading vocabulary test with the norms for the test is illustrated in Table 12. This table is a frequency distribution made from the test scores of the 28 pupils in this first grade class. The norms with which this

[8] This point is clearly demonstrated in Illustrations 7, 8, and 9 of the "Work-Book in Educational Measurements" by H. A. Greene. This *Work-Book* also furnishes extensive opportunity for drill on the finding the median from all types of distributions. Also see the *Appendix,* pages 364 to 369, particularly problems 11 and 12.

class was to be compared are medians, therefore, in order to make the comparison of this class with the norms perfectly fair, the class median was computed. The work of computing the median, step by step, is given in Table 12. The

Table 12

DISTRIBUTION OF SCORES ON A FIRST GRADE READING VOCABULARY TEST AT THE END OF 8 WEEKS. TEACHER, MISS W.

Class Interval	f
107.5 - 112.5	1
102.5 - 107.5	0
97.5 - 102.5	1
92.5 - 97.5	2
87.5 - 92.5	2
82.5 - 87.5	5
77.5 - 82.5	2
72.5 - 77.5	3
67.5 - 72.5	4
62.5 - 67.5	2
57.5 - 62.5	1
52.5 - 57.5	1
47.5 - 52.5	0
42.5 - 47.5	2
37.5 - 42.5	0
32.5 - 37.5	0
27.5 - 32.5	0
22.5 - 27.5	1
17.5 - 22.5	1
	N = 28

N = 28

$\frac{N}{2} = 14$ = Halfway

1+1+2+1+1+2+4 = 12 = Subtotal

14 − 12 = 2

$\frac{2}{3}$ of 5 = 3.3

72.5 + 3.3 = 75.5 = median

The median number of words in this test which my first grade pupils knew at the end of 4 weeks work was 22.0. This class has therefore gained approximately 54 words in the past four weeks of teaching. The test norm at the end of the first semester is 174 words. I must speed up my work somewhat.

student will do well to check this work carefully and see that all calculations are correct.

III. Measures of Variability

78. Need for Measures of Variability. The measures of central tendency represented by the arithmetic mean and the median are valuable statistical measures but they describe only one characteristic of the data, the tendency of the scores to pile up at or near the middle of the distribution. Descriptions of test results based wholly on one or the other of these measures are incomplete. Two classes may earn scores on a test which will have the same median, or mean, yet the instructional situations represented by the two classes may be decidedly different. This is due to the fact that the descriptions of the two classes lack any information as to the range of ability represented. This fact is illustrated graphically in Figure 6.

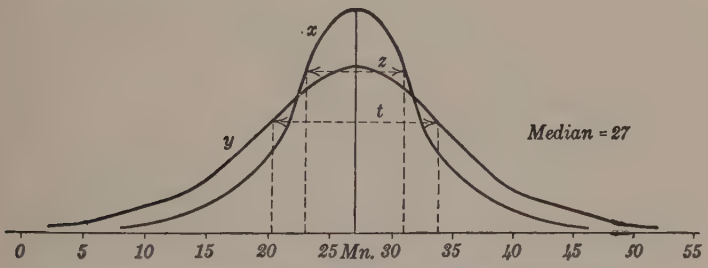


Figure 6. Illustrating Need for Measures of Variability.

It is a common practice to show frequency distributions in graphic form by representing the frequencies at a given point on the scale in terms of a line erected perpendicular to the base line or scale. If the tops of a large number of these perpendiculars are connected, the result is a curved line which usually is close to the base line at the ends of the scale but rises quite rapidly from the base line near the middle of the scale. In Figure 6 the curve marked x represents the distribution of scores made on a certain test by a

class. The curve being close to the base line at the ends or extremes means that there are relatively few very low and very high scores. The high point of the curve near the middle indicates that a great many pupils made scores near the average. This is typical of situations usually found where considerable numbers of cases are involved.

It will be noted in Figure 6 that while the middle portion of the curve x rises much higher than the similar portion of the curve y the extremes of curve x do not go out on the base lines in either direction as far as is true of curve y. This flatness or peakedness of the curve is the graphic indication of the variability of the data it represents. The less peaked the curve the greater the variability, other things being equal. It is thus apparent from this illustration that while the means of these two distributions are identical, very greatly different teaching problems are represented. Curve x represents a relatively homogeneous group while curve y is a more widely scattered group. If these two curves represent the distribution of ability of two classes, which one would you prefer to teach?

79. Range as an Expression of Variability. The *range* of scores, that is, the scale distance between the lowest and highest scores in a distribution is one way of expressing variability. However, it is one of the least reliable measures of variability or dispersion since it is apparent that it is affected by the fluctuation of the extreme scores. These extreme scores are very often the results of the operation of some unusual conditions. In the case of the scores reported in Table 6, page 119, the range is from 24 to 72, or 48 units. This range would be affected seriously if no pupil had made a score of less than 34. Such a chance difference in score would have affected the arithmetic mean only slightly and the median probably not at all. From this it must be clear that the range, because of the fact that it is dependent upon the extreme scores in a distribution, themselves subject to

the greatest variation, is an unsatisfactory measure of variability.

80. "Q" or Semi-inter-quartile Range as a Measure of Variability. A much more reliable expression of variation is found in the *semi-inter-quartile-range* which is really the range of the middle 50% of the cases. Thus the operation of the unusual deviation of "sport" cases at the extremes of arrays is eliminated. In actual practice *the range of the middle half of the cases is expressed in terms of one-half of the inter-quartile range and is called "Q."* The values of the upper quartile (Q_3) and of the lower quartile (Q_1) correspond respectively to the seventy-fifth percentile and the twenty-fifth percentile. The value Q and the median (Q_2) are distinct, however. One is a measure of variation or spread while the other is a measure of central tendency. The difference between the median (*a measure of central tendency*) and the quartile deviation (*Q, a measure of variability*) is made clear by Figure 6, page 139. The median line is the heavy vertical line labeled *Mn* which cuts the base line at the value 27 on the scale. The median is a *point* on the scale. The value Q is a *distance* on the scale, represented by the lines t or z in the diagram. The difference in the variability of the two distributions represented by the curves x and y is shown by the relative lengths of the lines t and z. In practice the value Q for the curve x is stated as one-half of the line z.

It has been pointed out that the central tendencies of the groups represented by the curves x and y in Figure 6 are identical while the spread of ability is much greater in distribution y than in x. Note that line t which represents the range of the middle 50% of the cases in distribution y takes in a much wider range of scores than is the case for line z which represents the range of the middle 50% of the scores in distribution x. All measures of variability tell us in one way or another the average or typical amount of deviation

from the average of the group as a whole. Central tendencies reveal the most likely point of balance between the high and low scores in a distribution. Measures of variability show how the group scatters from this central tendency. A distribution with a large measure of variability scatters more widely from the central tendency than does the distribution with a smaller measure of variability.

A comprehension of this general concept underlies an understanding of practically all of the measures of variability as well as a great many of the statistical techniques yet to be considered in connection with the interpretation of tests.

The quartile deviation (Q) and the standard deviation or sigma (σ) are representative of the more important measures of variability, and constitute the ones treated intensively in this book.

81. Computing the Quartile Deviation (Q). The method of computing the semi-inter-quartile range, or quartile deviation (Q) is demonstrated by the use of the data given in Table 13, page 143.

The first quartile or twenty-fifth percentile and the third quartile or the seventy-fifth percentile of a distribution are found in the same manner as is employed in finding the median except that, of course, the numbers of cases counted off above or below the points on the scale are different. For the first quartile one-fourth or 25% of the cases are counted off below the point taken as Q_1. For the third or upper quartile 75% of the cases are counted off below the point taken as Q_3. This point (Q_3) may be more easily obtained by counting off 25% of the cases from the top of the distribution.

The following statements of the specific steps necessary in the calculation of the quartile deviation (Q) are based upon the data given in Table 13.

Step 1. The third or upper quartile (Q_3) is a point on the

scale of this table such that 75% of the cases are below it and 25% are above it. Therefore count into the distribution from the upper end until one-fourth ($N/4$) of the cases are taken. Make the interpolation into the class-interval in exactly the same manner as is done in finding the median. In this illustration $N/4$ is 31.25 cases. There are 27 cases down to the top of the step beginning 92.5. This leaves 4.25 to be taken from the 9 cases in this step. The resulting interpolation involves taking 4.25/9 of a step of 5 which equals 2.36. This amount must be subtracted from the *upper* limit of the step (92.5) which gives a Q_3 of 90.14.

Table 13

Class Intervals	f
147.5–152.5	2
142.5–147.5	3
137.5–142.5	0
132.5–137.5	0
127.5–132.5	1
122.5–127.5	1
117.5–122.5	1
112.5–117.5	2
107.5–112.5	3
102.5–107.5	2
97.5–102.5	3
92.5– 97.5	9
87.5– 92.5	9
82.5– 87.5	8
77.5– 82.5	8
72.5– 77.5	12
67.5– 72.5	8
62.5– 67.5	6
57.5– 62.5	10
52.5– 57.5	8
47.5– 52.5	12
42.5– 47.5	8
37.5– 42.5	6
32.5– 37.5	3
	$N=125$

Step 2. Q_1 is found by counting into the distribution from the lower end one-fourth or 25% of the cases. In this illustration $Q_1 = 53.91$.

Step 3. The Semi-inter-quartile range or Q is found by taking one-half of the difference between the values of Q_1 and Q_3. The formula is $Q = \frac{Q_3 - Q_1}{2}$. In this illustration $Q = 18.11$, which means that for this distribution a distance of 18.11 units in both directions from the median will include one-half of the cases in the distribution when the distribution is approximately symmetrical or *normal*. Most educational data are sufficiently normal to permit this interpretation.

82. Further Practice in Finding Q. For opportunity for practice on the computation of the Quartile Deviation (Q)

the student is referred to pages 369 and 370 of the *Appendix* of this book, particularly to Problem 13, and to Problems 17 to 21 inclusive in the "Work-Book in Educational Measurements." Skill in the application of these statistical methods in the interpretation of your test results will be developed and maintained only through extensive practice on practical problems.

83. Standard Deviation (σ, sigma) as a Measure of Variability. Such a simple method of expressing the variation of a distribution of test scores as the Semi-inter-quartile-range (Q) is sufficient for most ordinary statistical situations but where careful work of an analytical or research type is being done a more exact means of expressing variability must be used. The standard deviation, also called sigma (σ), has many characteristics which make it a useful measure of variability. *The standard deviation is a sort of arithmetic mean of the deviations from the mean of the distribution.* It is a special type of mean of the deviations because of the method used in computing it. In calculating the standard deviation (σ) each individual deviation from the mean is squared, the sum of these squared values obtained, is then divided by the number of such deviations, and the square root of the result is then obtained. Restated, *the standard deviation (sigma) is the square root of the mean of the squares of the deviations from the mean of a distribution.* Expressed in symbols the standard deviation becomes $\sqrt{\frac{\Sigma fd^2}{N} - c^2)}s$ where Σfd^2 equals the deviations in the form of the sum of the products of the frequencies at each step by the deviation of each step from the assumed mean; N equals the number of cases in the distribution; c equals the correction as found in calculating the mean; and s equals the size of the class-interval of the distribution in units.

The relation of the standard deviation to the quartile deviation of a distribution is shown in the accompanying figure (Figure 7). By definition the *quartile deviation* (Q)

takes into account the middle fifty per cent of the cases. That is, the ordinates (lines erected perpendicular to the base line of the curve) erected at a distance equal to Q on either side of the mean or median include fifty per cent of the area of the surface between the curve and the base line. In the diagram the lines P and R represent the ordinates erected at a distance equal to Q on either side of the mean. The lines S and T represent ordinates erected at a distance equal to σ on either side of the mean. The standard deviation (σ) takes into account approximately 68 per cent (in

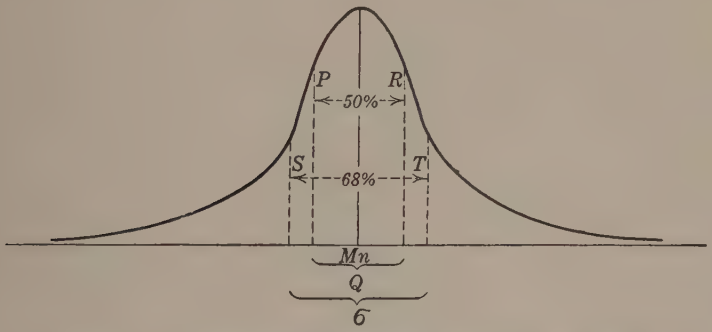


FIGURE 7. Comparison of Standard Deviation and Quartile Deviation (Q).

a normal distribution 68.26%) of the area of such a distribution. That is, ordinates erected at a distance equal to sigma on either side of the arithmetic mean include approximately two-thirds of the cases in the distribution.

The standard deviation may be computed about any measure of central tendency but it is common practice to consider it as computed about the arithmetic mean only. Unless otherwise indicated sigma may be considered as having been calculated about the arithmetic mean.

A further interesting characteristic of the standard deviation is indicated in Figure 7. Mathematically the value *sigma* bears a definite relationship to the curve of the distribution itself. It will be noted that where any large number of

cases or scores are found in a distribution there is a tendency for the larger portion of the cases to pile up in or near the middle of the distribution. When this type of distribution is presented in graphic form (as in Figure 7) the result tends to be a symmetrical bell-shaped curve with many cases in the middle and few cases at the extremes. Certain types of these characteristic bell-shaped distributions have come to be called *normal curves*, and for these normal curves mathematical formulae have been derived from which such typical curves may be computed if certain of the required data concerning the curve are given. In these formulae the value *sigma* is one of the quantities which must be given in order to construct such a curve from actual data. In this relationship the value sigma represents the distance from the mean at which the curve changes from convex to concave, a breaking point, as it were. In Figure 7 the points where the curve changes its character are indicated by the ordinates lettered *S* and *T*. The center of curvature for the convex portion of the curve is considered to be below the curve itself, and for the concave portion it is considered to be above the curve.

Thus, because of the relationship which the standard deviation bears to the curve of the distribution itself, and the reliable expression of variability which it provides due to the fact that every deviation in the distribution is considered, the standard deviation is one of the most useful of the measures of variability.

84. Computing the Standard Deviation (σ). According to the definition of the standard deviation it would seem that the arithmetic mean must be first computed. Where the data are ungrouped, as in the accompanying illustration, the mean must, of course, be found. When the data are in a frequency distribution it is not strictly necessary that the arithmetic mean be computed exactly although it is necessary to go through all but the last step of the process. The exact steps in the computation of the standard deviation from ungrouped data are given in detail in connection with Table 14.

Table 14

STANDARD DEVIATION FROM UNGROUPED DATA

Score	Deviation (d)	Deviations Squared (d^2)
95	7.6	57.76
93	5.6	31.36
90	2.6	6.76
89	1.6	2.56
87	.4	.16
86	1.4	1.96
85	2.4	5.76
85	2.4	5.76
82	5.4	29.16
82	5.4	29.16
874		$\Sigma d^2 = 170.40$

Mean = 87.4

$$\sigma = \sqrt{\frac{\Sigma d^2}{N}} = \sqrt{\frac{170.40}{N}} = \sqrt{17.04} = 4.13$$

The mean of the scores in Table 14 is 87.4. It is therefore clear that a score of 95 deviates from this mean by 7.6 points. A score of 93 deviates 5.6 points, etc. Since the standard deviation is a mean of the squares of these deviates from the mean of the array of scores, it is necessary to square each of these deviates. These are given under the column headed d^2. Since each deviation appears only once and the data are ungrouped the formula for the standard deviation can be simplified to read $\sigma = \sqrt{\frac{\Sigma d^2}{N}}$. In this illustration the sum of the deviations squared (Σd^2) is 170.40. The mean of these deviations is therefore 17.04 since there are exactly ten cases. The value 17.04 is the mean of the deviations

squared, therefore to turn it into units of the scale the square root of this quantity must be taken. This is found to be 4.13, which is the standard deviation (sigma) of this series of test scores. Thus we can think about the scores in Table 14 as follows: The central tendency (here, the mean) of this distribution is 87.4. The σ is 4.13. This means that between scores 4.13 points larger and 4.13 points smaller than this, approximately two-thirds of the scores are found.

Table 15

STANDARD DEVIATION FROM FREQUENCY DISTRIBUTION

Class Intervals	f	d	fd	fd^2 [9]
70.5–73.5	1	+6	+6	36
67.5–70.5	0	+5		
64.5–67.5	1	+4	+4	16
61.5–64.5	2	+3	+6	18
58.5–61.5	3	+2	+6	12
55.5–58.5	3	+1	+3 (+25)	3
52.5–55.5	6	0		
49.5–52.5	2	−1	−2	2
46.5–49.5	2	−2	−4	8
43.5–46.5	2	−3	−6	18
40.5–43.5	0	−4		
37.5–40.5	1	−5	−5	25
34.5–37.5	1	−6	−6	36
31.5–34.5	0	−7		
28.5–31.5	1	−8	−8 (−31)	64
	$N=25$			$\Sigma fd^2=238$

[9] The fd's represent the frequency of each step multiplied by the number of steps the interval lies away from the step in which the mean is located. The fd^2's are obtained by multiplying the fd's by d. This results in squaring each step-deviation.

85. Computing the Standard Deviation (σ) from Data in Frequency Distribution. The method of computing the standard deviation shown in the foregoing illustration may be applied almost directly to the calculation of the standard deviation from a frequency table. It must be remembered, however, that the simplified formula for the sigma does not hold here for when the data are in frequency tables, the deviation of the scores must be considered by groups and must be expressed in terms of steps in order to simplify the process. The accompanying problem (Table 15) which is completely worked through illustrates each step in the computation.

The formula for the standard deviation when data are in frequency distribution is $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - c^2}\, s$. The following calculations show all of the steps in the computation of the standard deviation of this distribution.

$$c = \frac{-31+25}{25} = \frac{-6}{25} = -.24;\ c^2 = .058$$

$$\sqrt{\frac{\Sigma fd^2}{N}} = \frac{238}{25} = 9.52 = S\,^{10}$$

$$S^2 - c^2 = \sigma^2$$
$$9.52 - .058 = 9.462$$
$$\sigma \text{ steps} = \sqrt{9.462} = 3.07\,^{11}$$

10 S^2 is used to designate the root-mean-square-deviation prior to application of the correction, c^2. The correction, c (squared to make it comparable to the S^2) is always deducted from the root-mean-square-deviation because this value, S^2, is *always too large*. If c has any value more than zero it means that the assumed mean and the true arithmetic mean do not coincide, therefore any deviations computed about it are too large and the correction c must be subtracted to compensate for this difference.

11 The student will find the tables of squares and square roots such as are given on pages 303–312 of Garrett's "Statistics in Psychology and Education" or Otis' "Statistical Methods in Educational Measure-

Size of step (s) = 3

σ units = 9.21 (Standard Deviation in Units of the Scale).

This work may also be expressed in words. In this problem the step with a mid-point of 54 is selected as the step in which the arithmetic mean is likely to fall. It is not necessary to compute the arithmetic mean but the correction must be found. The value of c is found to be —.24 which indicates that the assumed mean is somewhat larger than the true arithmetic mean of the distribution. In order to turn c into a form comparable to the other deviations used in the problem it must be squared. Thus c^2 equals .058. The sum of the squares of the deviations in the distribution (Σfd^2) is 238. The mean of these squares of the deviations (S^2) is 9.52. This quantity, 9.52, is too large by the amount of the square of the correction, c. The application of this correction gives the value 9.462 as the square of the standard deviation in terms of steps (σ^2 steps). Sigma in steps is therefore the square root of 9.462, or 3.07. To turn this into units 3.07 must be multiplied by the number of units per step in the distribution. The standard deviation (σ) is therefore 9.21. This means that if ordinates are erected at a distance of 9.21 units on either side of the arithmetic mean of this distribution approximately 68% of the distribution will be found between these ordinates. This will, of course, not be strictly true since no small frequency table such as the one used in this illustration will be sufficiently symmetrical to permit the cases to fall in exactly that fashion. Roughly this value, sigma, expresses the range of the middle 68% of the cases.

The arithmetic mean of the data used in this illustration is 53.28. The standard deviation (σ) in points on the scale is 9.21. Ordinates erected at 9.21 points above this arithmetic

ments,'' pages 299 to 305, very helpful in his work from this point on. Ability to use a slide rule or a mechanical calculator will also speed up the work as well as make it more accurate.

mean, and at 9.21 points below it fall at 44.07 and 62.49 on the test scale. Within these points will be found approximately 68% of the measures in the distribution when it represents enough cases to be approximately symmetrical.

86. Additional Practice in Finding Sigma (σ).[12] In Table 16 part of the work involved in computing the Standard Deviation of the distribution given has already been done in outline form. Certain of the steps have been left out to afford the student practice in making these computations. The answer to the problem is given as a guide in checking the work.

Table 16

SCORES OF A TENTH GRADE CLASS ON A PLANE GEOMETRY TEST

Class Intervals	f	d	fd	fd^2
97.5–102.5	1	6	6	36
92.5– 97.5	0	5	0	0
87.5– 92.5	1	4	4	16
82.5– 87.5	0	3	0	0
77.5– 82.5	1	2	2	4
72.5– 77.5	3	1	3	3
67.5– 72.5	6	0	(15)	
62.5– 67.5	7	−1	−7	7
57.5– 62.5	4	−2	−8	16
52.5– 57.5	2	−3	−6	18
47.5– 52.5	2	−4	−8	32
42.5– 47.5	0	−5	0	0
37.5– 42.5	1	−6	−6	36
	$N=28$		(-35)	$\Sigma fd^2=168$

Sigma (σ) $=\sqrt{\frac{\Sigma fd^2}{N}-c^2}\cdot s$ [13]

$c=\frac{-35+15}{28}=.714$

$c^2=.51$

$\frac{\Sigma fd^2}{N}=$

$\frac{\Sigma fd^2}{N}=S^2$

$S^2-c^2=$ $\quad=\sigma^2$

$\sqrt{\sigma^2}=$ $\quad=\sigma$ in steps

$s=5$ units. $\quad\sigma=11.70$

[12] Problems 19, 20, and 21 in Greene's "Work-Book in Educational Measurements" furnish useful practice in making these computations.

[13] s equals the size of the step in the distribution.

IV. The Relationship of Test Scores

87. Need for Measures of Relationship. The discussion up to this point in this chapter has concerned itself with two types or descriptions of test scores; namely (1) measures of central tendency, and (2) measures of variability. In connection with the more critical examination and interpretation of test results it often becomes necessary for the teacher or student to go somewhat beyond the description afforded by these simple measures. It is often necessary to select for testing purposes the one test of a series which most nearly measures the desired ability. The method followed in such a case involves finding the relationship between the several tests under consideration and the ability to be measured. This is called the method of correlation. It is the method applied when two or more measures of the same individuals are being studied with the view to finding the factors, elements or characteristics which are common to each. Sometimes the method of correlation is mistakenly used to attempt to discover causes operating to produce certain effects. There is nothing in the method or the result of computing a correlation which indicates definitely which of the factors is a cause and which is an effect, or whether both of the factors may be affected by other variables.

88. Meaning of Correlation. The degree of relationship between two series of test scores, or as they are frequently called, variables, may be indicated in a number of ways. The relationship may be shown graphically but for many who do not visualize readily this is an unsatisfactory method. It does not give a single brief means of describing the degree of relationship exactly. To meet this need a number of different methods of arriving at a single value which expresses this degree of relationship (correlation) have been devised, all of which have as their purpose the statement of the degree of relationship between the pairs of variables

in the form of a coefficient of correlation, called *r*. The possible values of *r* vary all the way from +1.0 indicating perfect positive relationship, through all of the possible decimals to zero (0) indicating no relationship whatever, to —1.0 indicating a perfect negative relationship.

The following are illustrations of a positive relationship between two factors:

(1) The rise and fall of a column of mercury in a thermometer with the rise and fall of the outside temperature. As the temperature rises the column of mercury also rises.
(2) The amount of the Sunday School collection and the number of pupils in attendance.
(3) The direction of the wind and the movement of smoke from a chimney. The smoke moves away with the wind.
(4) The tendency of pupils who are intelligent to be quite good readers also.

Negative correlation may be illustrated by:

(1) The movement of the elevator cage and the counter-balancing weights. As the elevator cage goes up the counter-balancing weights move in the opposite direction.
(2) The relation between absence from school and school achievement.

Zero or indifferent correlation is best illustrated by means of the chance matching of numbered cards which have been shuffled. Two packs of 25 blank cards each may be numbered and the packs carefully shuffled so that the cards are in no definite order. If cards are drawn from each pack and paired the resulting relationship is zero, or indifferent. If these same packs of cards are both arranged in ascending order and the first card from one paired with the first card from the other the resulting relationship is positive. If one pack is inverted and each time a small numbered card is taken from the one pack a large numbered card is taken from the other the resulting correlation is negative.

The illustration of the numbered cards used here suggests one of the simple methods of expressing correlation; viz.,

the method of ranking. If pupils are given two tests and the scores from the tests tend to place the same pupils in the same relative positions in each series there is an indication of a positive correlation between the two tests. For example, the accompanying pairs of scores (Table 17) earned by the nine pupils listed indicate a high positive relationship between the two tests because the pupil making the highest score on Test *A* also made the highest score on Test *B*, and each other pupil in the list maintained his relative position. This suggests that the two tests measure abilities which have a great many factors in common. On the other hand, if it had happened that Pupil No. 1 who made a score of 89 on Test *A* had made a score of 18 on Test *B*, and Pupil No. 2 who made a score of 85 had made a score of 20 on Test *B*, etc., the resulting correlation would have been negative, and would have suggested that the two tests were inversely related. That is, the negative correlation would indicate that the presence of ability on Test *A* implied the absence of ability on Test *B*.

Table 17

Pupil Number	Score on	
	Test *A*	Test *B*
1	89	32
2	85	31
3	83	29
4	80	28
5	76	26
6	70	24
7	65	21
8	61	20
9	54	18

There are formulae and methods by which the coefficient of correlation may be computed from data ranked as to relative position, but they are quite inexact and are not commonly used in the analysis of test results. The student who is interested in these methods will find them discussed in any of the textbooks on statistics referred to in this book.

It is probably useless for the student of educational measurements to study more than one method of computing the

coefficient of correlation. The one method we shall study, the Pearson Product-Moment method, is by far the most common and on the whole the basic method used in educational investigations. This method, while somewhat complicated and difficult because of the large number of different calculations to be made, really involves very little that the student has not already mastered. *This coefficient of correlation indicated by* **r** *is a single numerical index which expresses the extent to which the pairs of corresponding measures of two variables tend to deviate similarly from their respective arithmetic means.* It is computed from data arranged in frequency tables, but these frequency tables are of a slightly different form from that which has been considered in any of the problem work in this book thus far. The method of tabulating paired data is explained in the following section.

89. Tabulating Test Scores in Double Entry or Correlation Tables. The double entry or so-called correlation table is used in computing a coefficient of correlation by the Pearson Product-Moment method. It is also useful frequently when more than one test is given to the same group of subjects. The principles used in setting up the class-intervals in this double entry table are identical with those used in making a simple frequency table. In fact, the correlation table may be thought of as a simple frequency table on a vertical or y-axis which also has been extended in a horizontal direction to show the place of each frequency in the distribution on the x-axis as well. Thus each tabulation or check mark in a correlation table stands for the result of two measures of the same individual. Table 19 on page 159 is an illustration of this type of double entry or correlation table. This table is also used to illustrate the steps in the computation of the coefficient of correlation itself. Table 19 shows that one pupil who read with a comprehension score of between 7.5 and 10.5 also read with a rate score of between 59.5 and 66.5

as measured by this particular test, etc. The tendency of the frequencies to group themselves along the diagonal of the table is itself an indication of the correlation which exists between the two factors.

90. Steps in the Computation of the Pearson Product-Moment Coefficient of Correlation. The calculation of the Pearson Product-Moment coefficient of correlation involves the solution of the following formula: $r = \dfrac{\dfrac{\Sigma xy}{N} - c_x c_y}{\sigma_x \sigma_y}$ in which r is the coefficient of correlation, N the number of cases, σ_x the Standard Deviation (in steps) of the distribution on the X-axis, σ_y the Standard Deviation (in steps) of the distribution on the Y-axis, c_x the correction on the X-axis, c_y the correction on the Y-axis, and Σxy the sum of the products of the deviations of each measure from the central tendency of the X- and the Y-axes. The solution of the formula involves the following specific steps:

(1) Tabulation of paired scores in a double entry table.
(2) The computation of the Standard Deviation (σ) of each axis of the distribution leaving σ in terms of *steps*.
(3) The computation of the sum of the products of the deviations of each measure from the central tendency of both axes (known in the formula as Σxy).
(4) The division of the Σxy by the number of cases.
(5) The subtraction of the algebraic product of the corrections of each axis.
(6) The changing of the resulting fractional value into decimal form retaining its sign. If the sign is negative the relationship is inverse and if positive it is direct. The possible values of this decimal form range from plus 1.0 through the possible decimals to 0 and minus 1.0.

Of these steps, Numbers 1 and 2 have been explained and illustrated in previous discussion. The major new step is the determination of the sum of the xy products.

The name "product-moment method" implies the signifi-

cant feature of the process. The relationship itself takes into account the operation of forces (frequencies) at varying distances (deviation in steps) from the point of rotation (the mean). Since each measure assumes a position with regard to both axes the resulting moments [14] must take into account this fact. For example, if a child makes a high score on each of two tests tabulated in double entry form his position in the table will be well up in the upper right quadrant of the table.[15] The moment represented by his position is, however, the product of the distance his score lies from the mean on the X-axis and the distance his score lies from the mean on the Y-axis as well. In the illustration (Table 18) such a score might be shown by a frequency in the compartment made by the intersection of the deviations plus $4y$ and plus $3x$. The resulting xy product (moment) of such a case is 4 x 3 x 1 or 12. If two cases were found at that point the moment would be 24, etc. The moments or products in the upper right hand and lower left hand quadrants are always plus if the tabulation is made upward and toward the right. The products in the upper left hand and lower right hand quadrants are similarly always negative in sign. The Σxy is the algebraic sum of the plus and minus products. The sign of this product determines the sign of the correlation. The complete process of computing a product moment correlation is shown in Table 19, page 159.

91. Illustration of the Computation of the Pearson Correlation Coefficient. The data for this illustration represent the Comprehension and Rate scores of 76 pupils as measured by

[14] The reader will recall from his high school physics that the term *moment* is given to the force which tends to produce rotation in an unbalanced body. For example, a beam resting on a fulcrum is brought into balance when the moments of the forces operating on it are equal. A weight or a force of two units operating on a beam at a distance of one unit from the point where it is suspended is equalled in moment by a weight or force of one unit at a distance of two units from the point of suspension. Similarly in mathematics forces are brought into balance by equating their moments.

[15] Assuming tabulation upward and to the right.

a certain silent reading test (Table 19). The rate scores (the speed of reading) are tabulated along the horizontal (X) axis, and the comprehension scores are tabulated along the vertical (Y) axis. The student should study this illustration until he becomes familiar with the steps involved. Only a very few new processes are introduced. Attention is called to the fact that the columns headed d, fd, fd^2 are required in finding the standard deviations and have nothing to do with the values marked xy.

Table 18

X Axis

Y Axis										
								1		4
										3
			−						+	2
										1
										0
										−1
										−2
			+						−	−3
										−4
	−4	−3	−2	−1	0	1	2	3	4	

There are possibly a few points in the computation of the correlation coefficient which deserve a little further comment. The work of finding the Standard Deviations (Sigma) of each axis of the double entry table should give no difficulty. *It will be noted, however, that the sigmas in this illustration are not expressed in units of the scale but are left in terms of steps.* The reason for this lies in the fact that if the sigmas are turned into units of the scale each

Table 19

CORRELATION TABLE SHOWING RELATION OF RATE AND COMPREHENSION AS MEASURED BY A CERTAIN READING TEST

Rate—X

Comprehension—Y		66.5–59.5	73.5–66.5	80.5–73.5	87.5–80.5	94.5–87.5	101.5–94.5	108.5–101.5	115.5–108.5	122.5–115.5	f	d	fd	fd^2	xy
	52.5–55.5									2	2	+7	+14	98	42
	49.5–52.5										0	+6	0		
	46.5–49.5									4	4	+5	+20	100	60
	43.5–46.5								1	4	5	+4	+20	80	56
	40.5–43.5							4	1	2	7	+3	+21	63	36
	37.5–40.5							4	1	3	8	+2	+16	32	30
	34.5–37.5						5	2	1		8	+1	+8	8	4
	31.5–34.5						5	1		1	7	0	(+99)		
	28.5–31.5				4		3				7	−1	7	7	8
	25.5–28.5			2	4		1	1			8	−2	16	32	26
	22.5–25.5			4	5						9	−3	27	81	66
	19.5–22.5		3	1	1		1				6	−4	24	96	68
	16.5–19.5		1		1		1				3	−5	15	75	30
	13.5–16.5			1							1	−6	6	36	18
	10.5–13.5										0	−7	0	0	
	7.5–10.5	1									1	−8	8	64	40
	Total (f)	1	4	8	15	0	16	12	4	16	(76)		(−103)	(772)	(+484)
	d	−5	−4	−3	−2	−1	0	+1	+2	+3					
	fd	−5	−16	−24	−30	(−75)		+12	+8	+48	(+68)				
	fd^2	25	64	72	60			12	16	144	(393)				

$$c_x = \frac{-75+68}{N} = -.092 \qquad c_y = \frac{-103+99}{N} = -.053 \qquad c_x c_y = .005$$

$$S^2_x = \frac{393}{N} = 5.17 \qquad S^2_y = \frac{772}{N} = 10.16 \qquad \sigma_x \sigma_y = 7.22$$

$$\sigma_x = 2.27 \qquad \sigma_y = 3.18$$

$$\frac{\Sigma XY}{N} = +\frac{484}{76} = +6.37$$

$$r = \frac{\frac{\Sigma XY}{N} - c_x c_y}{\sigma x \sigma y} = \frac{6.37 - .005}{2.27 \times 3.18} = +.882$$

of the other expressions of deviation (the xy's) must also be turned into similar form. This would involve a considerable multiplication with a corresponding amount of division later, all of which is useless and only introduces additional opportunities for errors in computation. The value of the coefficient of correlation as a rule is not seriously affected by the size of the step taken in making the table. It is assumed, of course, that the class-intervals will be reasonable, but they need not be the same size on both axes.

Care must be taken to keep the sign of the corrections on each axis. If one correction is positive and the other negative the resulting $c_x c_y$ product is of course negative. This results in an algebraic subtraction of a negative quantity which serves to increase the numerator of the formula for r and accordingly to increase r itself.

The only other point which may cause the student any trouble is in determining the sum of the xy products (Σxy). Let us refer again to Table 19, page 159. The frequency *2* which appears in the intersections of the class-intervals 52.5 to 55.5 (Y) and 115.5 to 122.5 (X) lies above the arithmetic mean of the Y-axis seven (7) steps. At the same time these cases lie to the right of (above) the mean of the X-axis three (3) steps. The actual moment of these two cases is therefore the product of the distance they deviate from the mean of the one axis by the distance they deviate from the mean of the other axis ($2 \times +7 \times +3$ or $+42$). The sign of the deviation involved must be carefully noted. Let us take the 8 cases which lie in the class-interval 25.5 to 28.5 on the Y axis. Two of these cases deviate —2 steps from the mean of the one axis and —3 steps from the mean of the other. The product-moment of these two cases is therefore $2 \times -2 \times -3$ or $+12$. The product-moment of the 4 cases is $4 \times -2 \times -2$ or $+16$. The one case which deviates —2 steps from the mean of the one axis and 1 from the mean of the X-axis has a product-moment of $1 \times -2 \times 1$ or -2. The total of the plus moments (12 and 16) is $+28$. The resultant sum of the xy's for this particular row is therefore 28 — 2 or

+26. The other *xy's* in this illustration are found in a similar manner.

From this point in the calculation of the correlation coefficient the work is merely that of substituting values in the formula itself. The sum of the *xy* products (+484) is to be divided by the number of cases in the scatter diagram (76). The result is +6.37. From this must be subtracted the product of the corrections[16] of the *X*- and *Y*-axes ($c_x c_y$.) This reduces the numerator of the fraction represented by the formula for *r* to +6.365. The product of the sigmas of the *X*- and *Y*-axes is 7.22. The decimal represented by the fraction $\frac{6.365}{7.22}$ is +0.882, the coefficient of correlation (r) of the data in this illustration. In general, this coefficient of correlation indicates that there is a positive and rather significant relationship between the rate of reading of the 76 pupils tested and the comprehension with which they read.

92. Meaning of the Correlation Coefficient. The method of calculating the correlation coefficient as outlined in the foregoing pages is quite mechanical and, as such, can be mastered readily by most students. The interpretation of the meaning or significance of a correlation coefficient is often quite another matter for no entirely satisfactory mechanical device for accomplishing this has thus far been developed. A number of suggestions have been made recently, however, by means of which the student may be aided in attaching meaning to the correlation coefficient.

It will be remembered that correlation is usually indicated by means of what is called a *double-entry table,* or *correlation table.* The distribution of scores in such a table is known as a "scatter diagram." The appearance of the scatter diagram itself gives some indication of the amount of relationship which exists between the two variables shown. Assuming that the scatter diagram is made by tabulating up-

[16] This is the same correction used in finding the arithmetic mean.

ward and toward the right which is the best uniform practice, a high r is usually found where there is a very definite clustering of the cases along what would be the diagonal of the table. This means that the cases tend to be grouped somewhat systematically in a practically straight line running from the lower left-hand corner of the table to the upper right-hand corner of the table. This type of grouping is shown in Table 19 on page 159. As the cases scatter from the line of this diagonal, the correlation is reduced. If the cases are scattered over the table in a generally circular arrangement the resulting correlation will approximate zero. As the relationship changes from positive to negative the elliptical grouping of the cases takes place along a diagonal running from upper left to lower right. After some experience with "scatter diagrams" the student will soon come to have a definite feeling with regard to the probable magnitude of the correlation to expect.

PER CENT OF FORECASTING ACCURACY FOR SPECIFIC VALUES OF r

Coefficient of Correlation	Per Cent of Forecasting Efficiency
1.00	100
.99	86
.98	80
.95	69
.90	56
.866	50
.80	40
.75	34
.70	29
.65	24
.60	20
.50	13
.40	8
.30	5
.20	2
.10	½

One of the most important uses of correlation methods is to permit the use of known values in the forecasting of unknown values. The value of knowing the correlation of silent reading ability and progress through the school is to enable us to prophesy school progress from the results of tests revealing silent reading ability. The magnitude of the relationship found determines the accuracy with which the forecast can be made. A correlation coefficient of 1.00 means that a forecast based on one of the variables will be 100% correct. However, a correlation coefficient of .50 does not mean that the fore-

cast based on it will be 50% correct. Such is far from the case. The table of values on page 162 shows the relation between the indicated correlation coefficients and the per cent of accuracy of the forecast which may be based on the values.

This table is the result of applying Kelley's proposed formula for the *Coefficient of Alienation* ($k = \sqrt{1 - r^2}$) and then deducting this Alienation Coefficient expressed as a per cent from 100%. The Alienation Coefficient reveals the lack of relationship, and hence the inaccuracy of any forecast based upon the relationship. The complimentary value of $\sqrt{1 - r^2}$ expressed as a per cent thus represents the per cent of the forecasting efficiency of the correlation coefficient.

The accompanying brief summary of the significance of correlations of various sizes is adapted from a discussion by Clark L. Hull.[17]

Coefficient of Correlation:	Practical Interpretation
Below .50	Practically useless for forecasting purpose (only 13% effective).
From .50 to .60	Evidence of some relationship; possibly useful.
From .60 to .70	Evidence of a real relationship; of limited usefulness for prognostic purposes.
From .70 to .80	Evidence of a marked relationship; of rather marked prognostic value.
Above .80	Evidence of unusual relationship; marked prognostic value.

A further illustration of the practical meaning of correlations is given in the accompanying table of typical coefficients each of which is representative of a real type of school situation.

[17] Hull, Clark L., "The Correlation Coefficient and Its Prognostic Significance," *The Journal of Educational Research*, Vol. 15, May, 1927; pp. 327–338.

A Correlation Coefficient of	Is the Sort of Relationship Which Will on the Average be Found for:
$r = .94$	Two forms of a 90-minute high school content examination given to college students.
$r = .90$	Two forms of a 45-minute group mental test.
$r = .80$	The repetition two days apart of a certain 30-minute high school reading test.
$r = .70$	The average of three individual teachers' judgments of the reading ability of a class and the scores made by the class on a certain silent reading test for the elementary school grades.
$r = .60$	The judgments of a single teacher of the ability of a class of 6th grade pupils to solve verbal arithmetic problems with the class scores on a certain Problem Analysis test.
$r = .50$	The intelligence test scores and the class grades of a high school class in history.
$r = -.35$	(Negative relationship.) The chronological ages of a 7th grade class (or any other school grade for that matter) and the achievement scores made in a measurable school subject.

The student will learn to identify and utilize many other criteria and suggestions for the interpretation of correlations as he becomes more experienced in this type of work.

93. Additional Experience with Correlations. It is useless to expect that this elementary discussion of the methods of computing the correlation coefficient will be sufficient to give the student an adequate mastery without an opportunity for further drill. The student who is interested in going beyond a casual understanding of the general methods of correlation will wish to have an opportunity for practice on the calcula-

tion of correlation coefficients himself. Such an opportunity is afforded in the "Work-Book in Educational Measurements" [18] which is designed to give mastery over certain fundamental statistical skills needed in working with tests. Students interested in mastering the theory of correlation methods, or in applying the methods of more advanced research in the field of tests, are referred at once to the many excellent textbooks on statistical methods designed especially to serve that purpose.

Finally a word of warning should possibly be given. Do not let yourself become over-optimistic in the interpretation of correlations. Better be on the safe side when making claims for the reliability or the forecasting power of a test. Much irreparable damage has been done to the cause of educational measurements by the unqualified and exaggerated statements of misinformed individuals. This damage has taken two forms. As a result of frequent misinterpretation of the measures of relationship many highly ineffective tests have been accepted and used widely. Many users of educational tests have discovered for themselves the inefficiency of the devices recommended to them and have swung to the opposite extreme as a result of their awakening and will have nothing whatever to do with tests.

V. Assignment of Relative and Absolute Ranks

94. Need for Method of Ranking. It has been pointed out repeatedly in the discussion in foregoing sections of this book that test scores are given meaning only when they are considered in relation to some standard or norm of achievement. In working with test scores it often becomes desirable to make the achievement of other pupils in the group the basis for the comparison. The relative achievements of pupils may be compared by the simple process of assigning ranks or positions to their scores in accordance with their magnitude. That is, the pupil making the highest score is given

[18] See particularly Problems 22 to 25.

first position in the class, the pupil making next highest is given second position, etc. This is called the assignment of *relative rank*. It is a useful device for turning test scores which are otherwise not comparable into comparable form.

The importance of resorting to some device such as the ranking of scores to take care of the fact that scores from different tests or examinations are not comparable may be easily demonstrated. Any teacher readily realizes that a mark of 80 on an examination or a test may mean almost anything. It means one thing when there are no other marks of 80 given in the class. It means quite another thing when there are thirty marks of 80 given in the class. It may also be suggested that it means another thing depending upon how many pupils there are in the class. If the 80 is the highest grade or score made in a class of 50 pupils such a score is undoubtedly a superior mark. If the 80 is excelled by thirty pupils in the class of fifty pupils it means that the individual ranks 31st in a class of fifty. This point is discussed further in connection with the explanation of absolute ranks.

95. Assignment of Relative Ranks. *Relative ranks are positions assigned to pupils or scores in terms of the array of which they are a part.* The problem of ranking is merely that of assigning positions to the scores as arranged according to size. The largest score is given first position; the next largest, second; etc. The only difficulty involved appears in the case of two or more scores of the same size. The following illustration (Table 20) will make clear the method of handling this situation.

Pupil A, with a score of 46, the highest score when the measures are arranged in descending order of magnitude, is assigned first rank. Pupils B and C, both with scores of 44, would normally be assigned second and third places but it is impossible to assign second place or third place to one rather than the other, so average rank, or a rank of 2.5, is assigned to both. The relative numerical position of each

succeeding score is held, however. Pupil D, with a score of 41 being the fourth individual in the series, is assigned fourth place. The position assigned to the pupil with the lowest score should coincide with the number of cases in the series except in case the last scores are tied.

Table 20

RELATIVE RANKS

Pupil..	A	B	C	D	E	F	G	H	I	J	K	L
Score..	46	44	44	41	40	39	37	37	37	31	29	28
Rank..	1	2.5	2.5	4	5	6	8	8	8	10	11	12

Care should be taken in the ranking of individuals from scores in order to avoid errors in the assignment of positions. The practice of recording such scores on small cards which can be sorted and arranged in order will eliminate many mistakes. If the series runs into many cases it may be advisable to utilize the simple tabulation scheme shown in Table 21 as an aid in the assignment of positions. For purposes of illustration the data given in Table 6 on page 119 may be used very satisfactorily since they are not presented in any systematic order. It will be noted that Table 21 is nothing more than a double-entry frequency table with the *tens* represented along the vertical axis and *units* along the horizontal axis. A score of 72 is tabulated in the 70 row and in the 2 column. The balance of the 37 measures shown in the table are tabulated in a similar way. Such a tabulation of scores affords a basis for their arrangement in systematic order and at the same time reveals the cases of tied rank which must be taken into consideration in assigning positions in the series. The black face numbers in each compartment refer to the rank positions assigned to the various values.

It must be clear to the student that the assignment of

relative rank or position to pupils earning certain test scores actually covers up something of the true situation. For example, the differences in the magnitudes of scores are submerged by the assignment of ranks. To illustrate, in the case of the data used in Table 21, a difference of 3 points on the test scale (from 28 to 25) makes a difference of only one rank position (from 35th position to 36th). A difference of only one test point (from 72 to 71) also makes a difference of one rank position. This is a point which should be kept in mind when using the method of relative ranks. Thus ranking tells us that one pupil is ahead or behind another pupil but fails to tell how much in terms of actual test scores he is ahead or behind the other pupil.

Table 21

		Units									
		0	1	2	3	4	5	6	7	8	9
Tens	70	3 /	2 /	1 /							
	60	10.5 //			9 /	8 /		6.5 //		4.5 //	
	50	20 /		19 /	17.5 //	16 /	14.5 //		12.5 //		
	40	27 ///			25 /		24 /	22.5 //		21 /	
	30	33.5 //		32 /		31 /		30 /		29 /	
	20					37 /	36 /			35 /	

Problem 14 on page 371 of the *Appendix* of this book and Problem 26 of the "Work-Book in Educational Measure-

ments''[19] give additional experience with the assignment of relative ranks to test scores.

96. Assignment of Absolute Ranks. The usefulness of relative rank is somewhat limited by the fact that it takes no account of the actual level at which the accomplishment takes place. A person ranking 32 in a group of 35 has a very low relative rank in this group. However, if he ranked 32 in a group of 250 the significance of his accomplishment would be greatly changed. *Percentile ranks* take this point into account by reducing the ranking to a basis of 100 units. A percentile score of 75 means that for the measures under consideration the individual assigned that absolute rank made a score which exceeds 75% of the individuals of his group without regard to the number of cases it contains.

The student will recognize this 75th percentile as a measure with which he has already had some experience. The 75th percentile of a distribution is exactly the same thing as the third or upper quartile (Q_3). The same methods by which the 25th, 50th, or 75th percentiles are obtained may be applied to the computation of any designated percentile. It is often useful to determine the 10th, 20th, 30th, 40th percentiles. These are called *deciles,* and are computed in exactly the same manner as the median and the quartiles.

97. Computing the Percentiles (Deciles) in a Frequency Distribution. The calculation of the 10th, 20th, 30th, etc. percentiles involves exactly the same general procedure as that followed in computing the median. The only significant difference is, of course, in the number of cases used in counting into the frequency distribution. In computing the median (50th percentile) exactly one-half of the scores are counted off. In computing the 25th percentile (Q_1) one-fourth of the cases are considered. In computing the 10th percentile quite naturally only one-tenth of the cases are counted off. The interpolation into the proper step is exactly

[19] Page 91.

identical to the method followed in computing the median. Table 22 shows the general methods and results without detailed comment. Since the student has already done work of a similar nature no difficulty in understanding the correctness of every figure in the table should be found.

Table 22

COMPUTING DECILES

Class Intervals	f	Cumulative Frequency [20]		Percentile Scores
70.5–73.5	1	25	90% of 25 = 22.5	90th = 63.75
67.5–70.5	0	24	80% of 25 = 20.0	80th = 60.50
64.5–67.5	1	24	70% of 25 = 17.5	70th = 58.00
61.5–64.5	2	23	60% of 25 = 15.0	60th = 55.50
58.5–61.5	3	21	50% of 25 = 12.5	50th = 54.25
55.5–58.5	3	18	40% of 25 = 10.0	40th = 53.00
52.5–55.5	6	15	30% of 25 = 7.5	30th = 50.25
49.5–52.5	2	9	20% of 25 = 5.0	20th = 46.50
46.5–49.5	2	7	10% of 25 = 2.5	10th = 39.00
43.5–46.5	2	5		
40.5–43.5	0	3		
37.5–40.5	1	3		
34.5–37.5	1	2		
31.5–34.5	0	1		
28.5–31.5	1	1		
	$N = 25$			

The computations in Table 22 give the percentile scores reported in Table 23. The score assigned as the 100th percentile is the highest score obtained by any pupil on the test used in the table given as an illustration. In this case it is assumed to be 71. The percentile score assigned as zero

[20] The cumulative frequencies which are obtained by taking the cumulative total of the cases from the bottom of the distribution toward the top are given in this table for convenience in computing the necessary interpolations.

(0) is the lowest score made by any pupil on the test. In this illustration it is assumed as 30.

Table 23

Percentiles (Deciles)	Interpretation	Test Scores
100	Score equalled or excelled by no student...	71.00
90	Score equalled or excelled by 10% of the students.....	63.75
80	Score equalled or excelled by 20% of the students.....	60.50
70	Score equalled or excelled by 30% of the students.....	58.00
60	Score equalled or excelled by 40% of the students.....	55.50
50	Median—score equalled or excelled by 50% of the students.....	54.25
40	Score equalled or excelled by 60% of the students.....	53.00
30	Score equalled or excelled by 70% of the students.....	50.25
20	Score equalled or excelled by 80% of the students.....	46.50
10	Score equalled or excelled by 90% of the students.....	39.00
0	Score equalled or excelled by practically 100% of the cases.....	30.00

98. A Short Method of Computing Percentile Scores. The method of computing percentile scores illustrated in Table 22 on page 170, is satisfactory where comparisons are to be made with percentile norms in which the percentages of cases equalling or excelling a given score are given in the norms for the test. When it becomes necessary to determine the percentile score corresponding to a specific test score this method is not satisfactory because of the large number of calculations which must be made. Short methods of calculating the percentile scores corresponding to specific cases

in an array have been suggested by Thurstone, Otis, and others, and are so simple and useful that they deserve some attention on the part of the student of educational measurements. The method suggested by Thurstone is illustrated here.

99. Computation of the Percentile Score Corresponding to a Given Test Score in a Frequency Table. In the computation of the percentile score corresponding to a given test score the calculations are all made from the data in a frequency distribution in which the mid-points and limits of each class-interval are each carefully designated. The determination of the percentiles is based upon the cumulative addition of the result of dividing the number of cases in the distribution into 100. This is called the *rate*. This rate really expresses the number of percentile units represented by each score in the distribution. If the number of cases in the distribution should happen to be 100 the rate would, of course, be one. If the frequencies (numbers of scores) are less than 100 the rate is more than one. In Table 24 the number of cases is 80 and the rate is 1.25. This rate is added once for each frequency (single score) found in each class-interval. For example, in Table 24, 1.25 is added once for the one case found in the class-interval with the upper limit of 57.5, once more for the one case in the interval with the top limit of 67.5, etc. The cumulative addition of this rate for each of the 80 cases in the distribution accounts for 100% of the cases. It is assumed in adding these cases cumulatively that since all cases in a class-interval are considered the resulting percentile value should be assigned to the upper limit of the step, and a value half-way between the upper and lower limits of the step is assigned to the mid-point. The lower limit of the lowest step in the distribution is arbitrarily assigned a percentile value of zero, and the top limit of the top step is assigned a value of 100.

In the actual use of such a table for the assignment of the percentile scores corresponding to given test scores the per-

centile score nearest the actual test score is assigned.[21] For example, in using Table 24, a test score of 75 points would correspond to a percentile score of 8.1 on this test. A test score of 107 would be given a percentile score of approxi-

Table 24

SHORT METHOD OF COMPUTING PERCENTILES

Limits of Class Intervals				Percentile Ranks		
Lower	Mid-point	Upper	Frequency	Lower	Mid-point	Upper
117.5	120	122.5	1	98.75	99.40	100.0
112.5	115	117.5	3	95.00	96.88	98.75
107.5	110	112.5	4	90.00	92.50	95.00
102.5	105	107.5	7	81.25	83.63	90.00
97.5	100	102.5	12	66.25	73.75	81.25
92.5	95	97.5	15	47.50	56.88	66.25
87.5	90	92.5	14	30.00	38.75	47.50
82.5	85	87.5	10	17.50	23.75	30.00
77.5	80	82.5	6	10.00	13.75	17.50
72.5	75	77.5	3	6.25	8.13	10.00
67.5	70	72.5	3	2.50	4.38	6.25
62.5	65	67.5	1	1.25	1.88	2.50
57.5	60	62.5	0	1.25	1.25	1.25
52.5	55	57.5	1	0.00	.63	1.25
$N = 80$						
$\text{Rate} = \frac{100}{80} = 1.25$						

mately 89, etc. Problem 15 on page 371 of the *Appendix* gives practice on finding the percentile ranks corresponding to given test scores.

[21] It is, of course, obvious that tables of percentile scores must be prepared for each separate series of test scores to be assigned percentile scores.

VI. Summary

It is the purpose of this chapter to present to the beginning student in educational measurements an elementary discussion of the most useful statistical tools of measurement. Two of the three commonly used measures of central tendency are discussed and explained. Two of the more useful measures of variability are also explained and illustrated. Their practical usefulness is demonstrated in terms of the wide differences in the instructional problems involved in the teaching of groups of the same average ability, but of widely different variability. A brief discussion of the general meaning and method of the correlation technique is given. Practical contact with the problems of correlation is provided through the use of the supplementary material in the *Appendix* of this volume and in the "Work-Book in Educational Measurements."[22] A real mastery of the correlation technique is probably not essential to elementary work in measurements. It is important, of course, for those interested in advanced reading and research in the field. The methods of computing relative and absolute ranks from test scores are also discussed briefly in this chapter. These methods are especially useful to the teacher who is working with informal objective tests, or with other partially or non-standardized material.

There are numerous other statistical devices and methods which aid in interpreting the results of tests. It is not likely that the average classroom teacher will have need for all of these in her work with tests. Indeed, teachers are not often called upon to do more than to calculate median or mean scores after giving and scoring the tests. However, an interest in and a knowledge of the meaning of test results increases with a mastery of the statistical methods by which they may be interpreted. This mastery can be gained only through

[22] Greene, H. A. For practical problem work on the skills discussed in this chapter see Problems 1 to 28 inclusive.

use. The student or teacher who is interested in a more complete mastery of the principles of statistical methods than is given in this very greatly compressed chapter should study the references which follow.

Chapter VII takes up the discussion of the problems of interpreting the results of using tests in the classroom.

SELECTED REFERENCES

Garrett, H. E., "Statistics in Psychology and Education," Longmans, Green and Company, New York, 1926.

Greene, H. A., "Work-Book in Educational Measurements," Longmans, Green and Company, New York, 1928.

Hines, H. C., "A Guide to Educational Measurements," Houghton Mifflin Company, Boston, 1923, Chapters I–IV.

Monroe, W. S., "The Theory of Educational Measurements." Houghton Mifflin Company, Boston, 1923, Chapter XII.

Odell, C. W., "Educational Statistics." The Century Company, New York, 1925, Chapters I, III, IV.

Otis, Arthur S., "Statistical Method in Educational Measurement." World Book Company, Yonkers-on-Hudson, New York, 1925. Chapters II, III.

Ruch, G. M., and Stoddard, G. D., "Tests and Measurements in High School Instruction." World Book Company, Yonkers-on-Hudson, New York, 1927. Chapter XIX.

Rugg, H. O., "Statistical Methods Applied to Education." Houghton Mifflin Company, Boston, 1917. Chapters III–VI.

Symonds, Percival M., "Measurement in Secondary Education." The Macmillan Company, New York, 1927.

Thurstone, L. L., "Calculation and Interpretation of Percentile Scores." *Journal of Educational Research,* Vol. 6, No. 3: 225–235.

CHAPTER VII

INTERPRETING THE RESULTS OF TESTING

The purpose of this chapter is to present an elementary discussion of the following important points in the interpretation of test results:

a. The use of norms and standards in giving meaning to test scores.
b. The differences between norms and standards.
c. The major functions of grade norms and their use in the interpretation of test scores.
d. The importance of age norms in the interpretation of test scores.
e. The meaning and function of "derived scores."
f. The meaning and use of Intelligence Quotients and Educational Quotients in the interpretation of test scores.
g. The use of derived scores such as grade scores and percentiles in making comparable test scores from different tests.

I. Test Norms and Standards

100. Norms as Distinctive Features of Standardized Tests. Standardized tests are distinguished by the fact that they are uniformly accompanied by norms which permit an accurate interpretation of the meaning of their scores. Many informal objective tests meet all other criteria of objective measuring devices except this one, but on account of the lack of norms for the evaluation of achievement are limited in their usefulness. Norms furnish the teacher with definite information as to what has been achieved under ordinary schoolroom conditions. They, therefore, suggest what may be expected in the way of achievement under similar conditions.

The importance of norms in the reading of meaning into test scores is readily illustrated. Miss D., a third grade teacher, is interested in measuring the achievement of her class in silent reading. She has no standardized silent reading test at hand but understands how to construct the type of exercise she wishes to use for the measurement of the comprehension ability of her class. In the library of her room she finds in a set of supplementary readers some material which her pupils have not studied. Using this material as a basis she prepares a series of comprehension exercises which she uses in her class. She finds that the median comprehension score for her class is 34 exercises answered correctly and that the median rate of reading the material is 82 words per minute. Are these results good, bad, or indifferent?

A partial answer to this question may be secured by having these same articles read and these same exercises answered by another third grade class. This is done, and Miss D. finds that the median comprehension score for the other section of the third grade is 29 exercises, and the median rate of reading is 76 words per minute. This tells her that her class did better than one other third grade class. If these tests were given to a large enough selection of third grade pupils with the resulting median scores given above, these scores might come to be tentative norms. Let us assume that the achievement of Miss D.'s class is somewhat better than that of a large group of other similar pupils. We still wish to know how much better it is. Would the superiority of five points in comprehension for her class be a significant advantage? How near to typical fourth grade achievement is her class? These questions are answered by the process of securing grade norms of achievement for the tests. In this way the scores from a test may come to take on meaning.

101. Norms vs. Standards. At the outset it should be made clear that *norms* and *standards* are not exactly identical, although common and somewhat careless usage has tended to

make them appear the same. One reason for this confusion in the meaning of the terms is to be found in the use of the word, *standardizing,* for the process of establishing norms. Undoubtedly the use of the word *standard* itself as descriptive of certain objective tests has also contributed to the confusion. *Norms represent the actual levels of achievement of typical school children under controlled conditions.* If the reader will associate the word *norm* with *normal,* meaning average or typical he will aid himself in distinguishing between *norms,* which represent a definite central tendency, and the word *standard* as explained below. *Standards imply an ultimate goal to be achieved.* The standard of handwriting quality may be that of perfect legibility, while the norms (actual achievement of pupils) for quality are usually far below the standard. Standard accuracy in arithmetical computations is commonly placed at 100% accuracy. This undoubtedly is desirable although the typical or normal performance of school children is considerably below that point on practically all standard tests. This is illustrated in Table 25. This table is based upon a survey of Arithmetic ability

Table 25

NORMS AND STANDARDS OF ACCURACY ON COURTIS, SERIES B, ARITHMETIC TESTS. (Addition)

	Grade				
	4	5	6	7	8
Standard of Addition Accuracy	100%	100%	100%	100%	100%
Norm of Addition Accuracy [1]	43.1	60.4	68.9	64.2	67.2

[1] From an unpublished report of Bureau of Educational Research and Service, University of Iowa, Iowa City, on a mid-year survey of Arithmetic results as measured by *Courtis Series B Tests.*

in a large number of Iowa schools in 1927. The data show that the standard of addition accuracy for all grades is 100% of all addition examples attempted. The norms for addition accuracy are somewhat lower. The norm (median or typical accuracy of a great many pupils) of accuracy for fourth grade is 43.1%; for the fifth grade is 60.4%; for the sixth grade is 68.9%; etc. This shows clearly that as far as these pupils are representative the norm of accuracy is considerably below the standard.

It should possibly be pointed out here that standards may represent 100% of perfection, or 75% of perfection, or any other level of accomplishment depending upon the decisions of those responsible for the educational situation concerned.

102. Uses of Test Norms and Standards. It has been pointed out in a foregoing section of this chapter that norms are distinctive features of standard tests. The distinctions between norms and standards as such have been indicated. The prevalence of norms rather than standards for most educational tests, however, forces us in the interests of consistency to use the term norms most frequently in this discussion. It should be remembered that many times a norm (an average or typical achievement) is not a respectable total of achievement. Ambitious teachers will recognize this limitation of test norms as usually interpreted and encourage their pupils to strive for a superior type of accomplishment.

A number of specific uses of test norms and standards are listed below:

1. It is by means of the norms and standards that test scores are given meaning. This is illustrated concretely by the data on Table 28 and 32 on pages 186 and 190.
2. Norms afford the basis for the evaluation of instruction. If the achievement of given school groups is consistently below the norm an investigation into the underlying causes should be made. See Figure 9, page 186.
3. Norms provide goals of achievement which are valuable for the stimulation of pupil effort.
4. Norms provide the basis for the analysis of individual

pupil's test scores for the purpose of revealing individual difficulties and weaknesses. See particularly Tables 33, 34, and 35 on pages 192 and 193.

5. Norms afford the basis for determining whether a pupil is working up to the level of ability which may be expected of him. See Table 39, page 200.
6. Reliable norms furnish the necessary information for the more accurate and objective classification and gradation of the pupil. See Table 37, page 196.

103. Kinds of Norms. The use which is to be made of any given test is one of the most important factors determining the types of norms which should accompany it. Norms are usually the median or average achievement of given groups of otherwise unselected school pupils. It has been suggested by certain test authors that a goal of achievement somewhat higher than that represented by the mean or median should be used. The 60th or the 70th percentile has been proposed as suitable. These represent standards of achievement rather than norms.

The manner of grouping the pupils used in deriving the norms determines the kind of norm provided. If the pupils are grouped by school grades without regard to their ages or the length of time they have been in school the resulting norms are *grade norms*. If they are grouped by ages, without regard to the grades in which they are found the resulting norms are *age norms*. Norms are also being established to show the differences in achievement which may be expected from pupils who are normal, retarded, or accelerated [2] for their grades. These are progress norms.

104. Derivation of Norms. The grade norms established for most of the commonly used achievement tests are based on

[2] Normal children are pupils who have received the expected promotion into the succeeding grade each year they have been in school. That is, a child who has been in school five years and has made five years of school progress is *normal*. If he has been in school more years than he has made grades of school progress he is said to be *retarded*. If he has been able to make more than one grade of school progress for any given year of school attendance he is said to be *accelerated*.

the median scores obtained by giving the tests to thousands of pupils in each grade. Such norms provide the basis for the interpretation of class scores as well as of individual

Table 26

DISTRIBUTION OF A SAMPLING OF TOTAL SCORES ON COMPASS SURVEY TESTS IN ARITHMETIC
(Elementary Examination)

Mid-point of Class-Intervals	Grade			Mid-point of Class-Intervals	Grade		
	2	3	4		2	3	4
114	.....	.....	7	48	17	16	3
111	1	1	22	45	30	16	
108	.....	4	46	42	24	7	1
105	.....	12	50	39	40	9	2
102	.....	21	44	36	39	10	1
99	.....	24	43	33	38	3	
96	.....	17	33	30	40	3	
93	1	12	24	27	34	5	
90	.....	20	20	24	32	2	
87	.....	29	17	21	23		
84	.....	27	15	18	18	1	
81	1	26	6	15	17		
78	3	30	5	12	12	2	
75	1	18	5	9	9		
72	.....	26	7	6	5	.....	1
69	1	27	2	3	2		
66	2	33	6	0	1		
63	2	25	3				
60	7	20	3	Total.....	437	506	371
57	9	28	2				
54	7	16		Median...	33.51	72.11	99.35
51	21	16	3				

accomplishment. They are commonly considered as defining the degree of ability which a pupil or a group of pupils should possess. In some instances the idea of perfection

is associated with a norm. However, present methods of standardizing a test should make clear that we are not justified in thinking of norms as statements of what ought to be. As they are now derived, norms should be interpreted as representing the average achievement of certain groups of pupils under present school conditions.

The derivation of grade norms is accomplished by the simple process of giving the test to be standardized to a large and unselected group of pupils in the grades for which the norms are desired under the conditions which are to be maintained in the later use of the tests. The actual steps in the process are illustrated in Table 26. This table represents the frequency distributions of the scores made by the

Table 27 [3]

PERCENTILE NORMS SMOOTHED TO THE NEAREST WHOLE NUMBER
(Based on Data Given in Table 26)

Percentile Scores	Grade		
	2	3	4
99	79	106	114
90	51	99	109
80	45	90	107
75	43	87	105
70	41	84	104
60	37	79	102
50	34	72	99
40	30	67	97
30	27	61	93
25	25	58	90
20	24	55	87
10	16	45	75
1	5	26	40

[3] A further illustration of the steps in deriving grade norms for test material is given in Problem 29 of the "Work-Book in Educational Measurements," by H. A. Greene.

pupils in grades 2, 3, and 4 for which the norms are desired. The data are taken from a few of the actual cases used in deriving the norms for Total Score on the *Compass Survey Tests in Arithmetic.*[4] In this case the norms are arranged to show not only median achievement per grade, but also each decile of achievement (i.e., each tenth percentile, as 10th, 20th, 30th percentile, etc.). The norms themselves based on these distributions (Table 26) are shown in Table 27, on page 182. Figure 8, page 184, shows in graphic form the median, 25th percentile, and 75th percentile of achievement for this test.

The steps in the process of deriving grade norms for a test represent in a general way the procedure involved in securing age norms, or progress norms. Percentile norms are not commonly reported for the different age groups. They afford very helpful supplementary information for the interpretation of grade norms, however.

105. Meaning of Test Norms. In the use of standard tests for general survey purposes or for purposes of group comparisons grade norms ordinarily afford the chief basis for interpretation. The meaning which may be attached to such interpretations, however, depends upon a number of factors which the norms themselves do not take into account directly. For example, it is assumed that when tests are being given for the purpose of deriving norms of achievement the exact conditions operate which are to be maintained when the tests are given in the classroom. The subjects used are grouped according to the training they have had in the field of the test, or according to the experience they have had in the taking of tests. Certain time limits of work are established. Certain statistical procedures are utilized in analyzing the data. If comparisons with norms affected by these conditions are to be valid the results compared must be secured under identical conditions. The specific directions accom-

[4] *Compass Survey Tests in Arithmetic,* devised by Greene, H. A., Knight, F. B., Ruch, G. M., and Studebaker, J. W. Published by Scott, Foresman and Company.

panying standard tests must be followed exactly if comparisons with norms are to be made. The groups of subjects

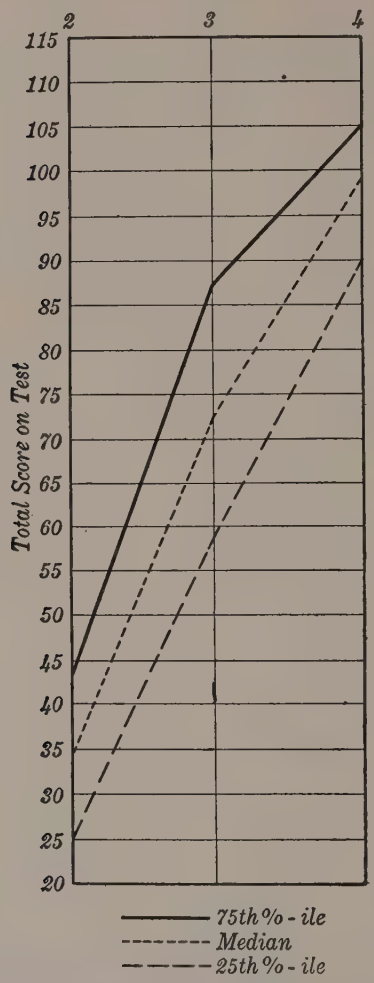


FIGURE 8. Graphic Presentation of the Median, 25th, and 75th Percentiles given in Table 27.

(pupils) must themselves be comparable in terms of training, experience, and criteria of classification. Statistical tech-

niques similar to those used in setting up the norms must be employed in making the comparison. That is, if the median achievement of the standardization group is used as a norm then the median scores of the classes to be compared with the norms should be used. If a percentile score such as the 60th percentile is used as the norm, then a similar percentile should be used for comparison. If the norms are established for a particular period or amount of school training, account of this must also be taken in making any interpretations of results.

Since norms reflect average or typical achievement it is, of course, obvious to the student that the school system which is above the average in such matters as the social and mental level of the population from which it draws its pupils, the quality of the physical equipment, the training and qualifications of its teachers, etc. should not be satisfied with merely normal scores. Norms are for the guidance of the average. Those who wish to excel in their achievement should establish higher goals for themselves.

106. Grade Norms. The function of grade norms for group comparisons may be illustrated by data from the *Woody-McCall Mixed Fundamentals Tests in Arithmetic.* The solid line in Figure 9 represents the achievement which may be expected from typical children in grades three to eight inclusive. The grade medians for these grades in a certain school system tested in October are shown by the broken line. This comparison shows that with the exception of the third grade and seventh grade the achievement of the pupils in this school system on this test is practically one-half a year or grade behind that of similar pupils in other school systems. The data upon which this comparison is based are given in Table 28.

These test norms, which are taken from the class record sheet for the tests afford a nice illustration of an attempt to adjust the norms to meet differences in extent of school training. Users of these tests are directed by the authors

to add the amounts shown in Table 29 to each grade standard for each month after October in order to make the norms provided comparable with results obtained later in the school year.

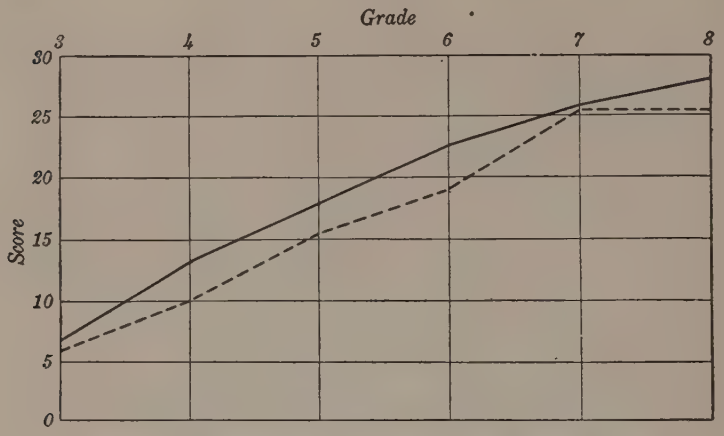


FIGURE 9. Comparing Achievement with Grade Norms.

Table 28

WOODY–MCCALL TEST RESULTS

Grade	3	4	5	6	7	8
Norm	6.8	13.1	17.8	22.5	25.9	27.8
School A (October)	6.1	9.9	15.2	19.0	25.4	25.5

Table 29

Grade	3	4	5	6	7	8
Increment	.54	.43	.42	.24	.25	.20

The increment .54 for the third grade means that for each month of school training or experience third grade pupils will increase their achievement on this test an average of .54 test points. Thus, if the third grade norm is 6.8 (Table 28) the norm for this grade for November should be considered as .54 test points higher, or 7.34. If the norm for the fourth grade is 13.1 for October, the November norm should be .43 points higher or 13.53, etc. This is a practice which should probably be adopted more generally in connection with grade norms for tests. In a certain sense it means the setting up of norms of growth or improvement; a thing which has considerable value for the teacher but which has not yet been done extensively. In connection with a state testing program in silent reading, norms for semester gains in comprehension based upon end-of-year results on Form 2 compared with middle-of-year results on Form 1 of the *Monroe Standardized Silent Reading Tests* were reported.[5] These improvement norms are probably fairly reliable since they are based upon test returns from 18,000 identical cases of pupils tested at the beginning and end of the semester. The results of this survey showing the median amount of improvement in silent reading comprehension as measured by the *Monroe Silent Reading Tests* are given in Table 30. The semester improvement is expressed in terms of the differences between the grade medians for Forms 1 and 2 of the tests given in January and May respectively. The Form 2 scores were changed to equivalent Form 1 values by means of a formula provided by the author of the tests. If tables of this type were available for many of the other measurable subject matter fields much good might come of it. It is assumed that the teaching process should result in bringing about changes in the responses of our students. It is also assumed that objective tests at least partially reveal this change. It seems logical to conclude that the efficiency of our teaching will finally come to be

[5] Greene, H. A., "A State Survey of Silent Reading in Iowa." Bulletin of Extension Division No. 91. University of Iowa, June 1, 1923.

evaluated in terms of the extent to which the expected improvement is actually brought about.

Table 30

SEMESTER IMPROVEMENT, MONROE SILENT READING TESTS [6]
(State-wide Grade Medians, Iowa Survey)

Grade	Form 1 (January)	Form 2 [7] (May)	Improvement
3	6.3	9.8	3.5
4	12.4	13.5	1.1
5	16.8	19.1	2.3
6	19.9	20.9	1.0
7	23.1	25.2	2.1
8	25.1	29.2	4.1
9	26.1	27.8	1.7
10	28.4	29.7	1.3
11	31.5	32.5	1.0
12	34.2	35.8	1.6

[6] *Ibid.*

[7] Form 2 changed to equivalent Form 1 value by Monroe's formula. This change in the Form 2 scores is necessary in order to make them directly comparable to Form 1 scores.

107. Age Norms. The inadequacy of the grade norms as a basis for comparison with the test scores of pupils is nicely illustrated by the data in Table 31. This table presents data taken from an unpublished study by Van Wagenen (1917) and is based on the scores of 476 fourth grade pupils in the schools of St. Paul. The scores of these fourth grade pupils are distributed by ages within the grade. The medians reveal a fact which many teachers have observed, namely, that the older pupils within a given grade make an average score which is lower than that made by the younger pupils within the same grade. This same situation is revealed graphically in Figure 10. The solid line represents the median scores made by each age group within the fourth grade. The broken horizontal line corresponds to the median score given under

the total column in Table 31. This figure shows very definitely how the median score for this fourth grade class would work an obvious injustice on the younger and more able members of the grade. A recognition of the evident inability of grade norms to cope with this situation has prompted authors of tests to provide norms for age groups as well as grade groups.

Table 31

DISTRIBUTION OF LANGUAGE TEST SCORES BY AGES
(476 Fourth Grade Pupils) [8]

Scores	Ages					
	8	9	10	11	12	Total
16–16.9	.....	2	.....	.....	.....	2
15–15.9	1	1	.....	.....	.....	2
14–14.9	1	6	.....	.....	.....	7
13–13.9	2	15	3	1	.....	21
12–12.9	2	14	9	4	7	36
11–11.9	6	28	18	2	1	55
10–10.9	8	40	30	8	6	92
9– 9.9	1	20	12	4	1	38
8– 8.9	5	40	33	18	6	102
7– 7.9	3	5	11	2	6	27
6– 6.9	.....	21	12	6	8	47
5– 5.9	2	5	4	1	2	14
4– 4.9	4	6	6	4	2	22
3– 3.9	1	.....	.....	1	.....	2
2– 2.9	1	2	3	1	1	8
1– 1.9						
0– 0.9	.....	.....	.....	1	.....	1
Total....	37	205	141	53	40	476
Median...	10.2	10.1	9.1	8.6	8.2	9.4

[8] From unpublished study by Van Wagenen.

Age norms are particularly valuable where it is desirable to make use of the test results of individual pupils rather

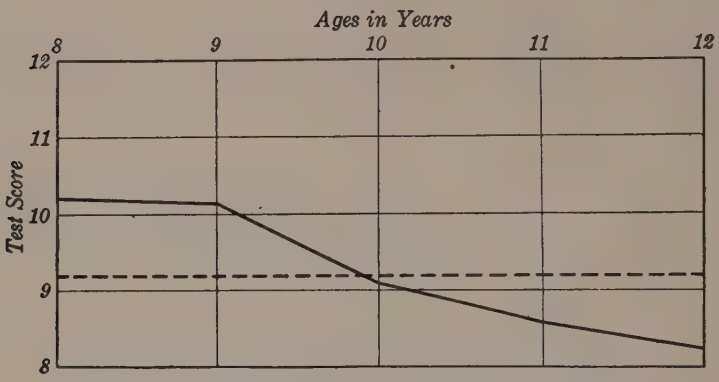


FIGURE 10. Illustration of Need for Age Norms.

than of entire classes or grades. The following illustration of such age equivalents (Table 32) is based upon a table

Table 32

AGE-EQUIVALENTS OF COMPOSITE SCORES OF ADVANCED EXAMINATION STANFORD ACHIEVEMENT TESTS

Composite Scores	Age Equivalents Yr.	Mo.	Composite Scores	Age Equivalents Yr.	Mo.	Composite Scores	Age Equivalents Yr.	Mo.
6	7	6	44	11	6	75	15	0
9	8	0	49	12	0	79	15	6
12	8	6	53	12	6	83	16	0
17	9	0	57–58	13	0	87	16	6
22	9	6	62	13	6	91	17	0
27–28	10	0	66–67	14	0	95	17	6
33	10	6	71	14	6	98–99	18	0
38–39	11	0						

accompanying the *Stanford Achievement Tests.*[9] For each half-year of age the corresponding composite score on the advanced examination is given. In the Stanford manual itself a table of age norms is also given for each of the separate parts comprising the battery of examinations. On the basis of the above norms the achievement age of a pupil may be obtained without regard to his grade classification. For example, suppose a pupil 11 years and 3 months of age makes a composite score of 53 on the advanced examination. A reference to this table immediately reveals the fact that he has made a score which is typical of a large group of pupils 12 years and 6 months of age. This age, 12 years and 6 months, thus becomes the educational or achievement age of this pupil. In view of the fact that the pupil's actual age is only 11 years and 3 months, this means that his achievement is one year and three months above the typical score of pupils of his age. Under the heading of "Derived Scores" (page 194), the method of turning these age scores into the form of quotients is discussed.

108. Interpretation of Grade and Age Norms. The interpretation of results from diagnostic tests for both individual pupil and class diagnosis may also be made direct from grade and age norms. This is illustrated by data from the *Compass Diagnostic Tests in Arithmetic* (Table 33). A fifth grade pupil R. H., age 11 years and 3 months, made the scores on the five parts of *Test 7, Multiplication of Fractions and Mixed Numbers* shown in Table 33 on page 192.

The age equivalent of 10 years and 9 months means that a total score of 75 points on this test is the typical or average score made by pupils who were 10 years and 9 months of age. See Table 35.

The symbol H6 means that a score of 11 points on Part 3

[9] *Stanford Achievement Tests, Manual of Directions,* Kelley, T. L., Ruch, G. M., and Terman, L. M. World Book Company.

of this is equivalent to the typical score made by pupils in the advanced section of the 6th grade. See Table 34.

Table 33

PUPIL RECORD FROM COMPASS DIAGNOSTIC TEST NO. 7

Part	1	2	3	4	5	Total
Score	17	13	11	30	4	75
Age Equivalent						10 years. 9 mo.
Grade Equivalent	H6	H6	H6	−5	−5	

The symbol —5 means that a score of 30 points on Part 4 of this test is less than available norms for the 5th grade. See Table 34.

Table 34

GRADE NORMS FOR COMPASS DIAGNOSTIC TEST NO. 7

Part	Grades						
	H5	L6	H6	L7	H7	L8	H8
1	16	16	17	17	17	17	18
2	12	12	13	13	14	14	15
3	9	10	11	11	12	12	12
4	35	42	49	53	57	62	67
5	8	9	10	11	12	12	13
Total	80	89	100	105	112	117	125

The age equivalents and grade equivalents which also appear in the table are obtained from the accompanying tables

or norms for the test (Tables 34 and 35). For example, a score of 17 on Part 1 of Test 7, *Changing Mixed Numbers to Improper Fractions,* is the typical or normal score for pupils in the advanced sections of the sixth grade. This pupil's achievement in *Cancellation in the Multiplication of Fractions* (Part 2) and in *Reducing Answers to Best Form* (Part 3) is also equal to that of high sixth grade pupils. However, in the knowledge of the *Fundamentals of Multiplication of Fractions* (Part 4) and in *Finding Errors* (Part 5) his achievement is considerably below that of the fifth grade norms. For purposes of individual diagnosis this is the significant information. This pupil is still in need of further practice on the multiplication of fractions. Possibly his lack of experience with exercises calling for the finding of errors may account for his relatively low score on that function.

Table 35

AGE-EQUIVALENTS OF TOTAL SCORES

Compass Diagnostic Test No. 7

Scores	Age Equivalents Yrs.	Mos.
66–70	10	6
71–75	10	9
76–79	11	0
80–84	11	3
85–89	11	6
90–94	11	9
95–98	12	0
99–102	12	3
103–105	12	6
106–109	12	9
110–112	13	0
113–115	13	3
116–117	13	6
118–120	13	9
121–124	14	0
125–128	14	3
129	14	6

A real understanding of the meaning and use of test norms in the interpretation of test scores can come only as a result of extensive first hand experience with their use. The "Work-Book in Educational Measurements" which was designed to parallel this and other textbooks in measurements with practical problems in the interpretation of tests furnishes this opportunity for experience. Problem 29 in the *Work-Book* shows the steps in the process of deriving a set of grade norms for a series of practice exercises

in arithmetic. Problems 34 to 38 inclusive in the *Work-Book* give further practice on the summarizing and interpreting of test scores from various types of tests. The student who does not have access to this *Work-Book* will do well to refer to Problems 16 and 17 on pages 372 and 373 of the *Appendix* to this volume.

II. Types of Derived Scores

109. Function of Derived Scores. Test scores, which are used to describe the performance of pupils are expressed in a variety of different units and in relation to a variety of different scales of measurement. In some tests the scale of measurement or the unit of measurement used may be relatively large. Pupil scores expressed in terms of these large units may be small. A long test composed of many exercises may result in large scores. Some common basis must be utilized if comparisons of scores based on these widely different types of scales are to be possible. This is illustrated in Table 36.

Table 36

Comparison of Units of Measurement Used in Two Reading Tests in Terms of Grade Norms

(*Thorndike–McCall* and *Iowa High School* Reading)

Test	Grade					
	7	8	9	10	11	12
Thorndike–McCall T-score	56	59.6	61.5	62.9	64.5	66.8
Iowa H. S. Silent Reading Test Total Comprehension Score, Tests 1–6	92	105	124	143	161	173

Table 36 presents a comparison of the grade norms for two silent reading tests both of which are standardized for use in grades 7 to 12. On the *Thorndike-McCall Silent Reading Scales* a T-Score of 56 constitutes the normal seventh grade score. Contrast with this a score of 92 as normal for seventh grade pupils on the *Iowa High School Silent Reading Test*. The comparison is even more striking when the normal scores for the twelfth grade achievement are compared. It is thus apparent that a score such as 90 on one test may mean exceptional achievement while a similar point score on another test may mean exceedingly poor achievement for the grade. To repeat, if scores on different tests are to be compared directly, they must be reduced to a common denominator. That is, if a score of 62 points on one test is poor performance but on a second test because of a different system of scoring is very good performance, we must have some way of translating both of these scores of 62 points to a common system of scores so that their difference in meaning will be clear. A number of proposals have been made for the calculation of derived scores which will partially take care of this difficulty.

110. Grade Norms as Derived Scores—Type 1. Point scores yielded by a test may be expressed in terms of the grade norms for the test. This is a very effective method of changing a test score into form which is readily understood, and which at the same time expresses the level of achievement represented by the score. These grade equivalents can be found for scores on any test having reliable grade norms. Their significance is, of course, subject to the reliability of the test norms as well as to the reliability of the test itself. Following its somewhat general use the grade equivalent has been featured in connection with certain tests for which age norms were not available. These "G-scores" as McCall has suggested calling them are based directly upon the grade norms for the tests, and may be computed quite readily for any test. The method is illustrated in Table 37. This table

shows, for example, that a point score of 56 is typical eighth grade achievement while a score of 52 is typical seventh grade achievement. Each point value from 52 to 56 is assigned its proportionate position on the grade scale. A point score of 53 is considered as equivalent.to the 7.3 grade, 54 to

Table 37

GRADE LEVELS FOR *Thorndike-McCall* SILENT READING SCALES

Score	Grade	Score	Grade
67	12.7	48	6.1
66	12.2	47	5.9
65	11.8	46	5.6
64	11.4	45	5.4
63	10.9	44	5.2
62	10.5	43	4.9
61	10.1	42	4.7
60	9.7	41	4.5
59	9.2	40	4.2
58	8.8	39	4.0
57	8.4	38	3.8
56	8.0	37	3.5
55	7.8	36	3.3
54	7.5	35	3.1
53	7.3	34	2.8
52	7.0	33	2.6
51	6.8	32	2.4
50	6.6	31	2.1
49	6.3	30	1.9

7.5 grade, etc. In other words, a score of 54 is such a score as may be expected from pupils half-way through the seventh grade. Table 37 shows the grade scores which may be assigned to the corresponding point scores on the *Thorndike-McCall Silent Reading Scale.*

Grade scores of this type are easily constructed from the grade norms ordinarily accompanying standard tests. See

Problem 30 of the "Work-Book in Educational Measurements." The degree of fineness with which this scale may be developed depends upon the characteristics of the test scale itself. If the test scale consists of many fine units, the grade scores assigned can be correspondingly minute. For example, a test with a relatively coarse unit requiring only one or two such units to span the entire difference between two grades would, of course, permit the establishment of the grade scores on a similarly crude basis. On the whole there is probably an element of danger in the general use of G-scores due to the fact that they tend to foster a feeling of mathematical exactness which does not exist. It is very doubtful if there is a single test in existence for which the G-scores for an individual pupil have meaning. The significance of such scores is entirely dependent upon the reliability of the test and of the norms from which they are developed.

This method of expressing test scores is particularly useful where test results are used for purposes of gradation and classification of pupils. The chief convenience lies in the fact that it expresses the achievement of the pupil in terms of the school unit of classification. The manner in which the test scores of a pupil on a number of standardized tests can be shown graphically by means of these grade equivalents is illustrated in Figure 1, page 57. This figure is representative of a simple type of test record card. It not only reveals the achievement of the pupil on the tests, but it also reveals at a glance the variation of ability in the different tests, and the accuracy of his classification. Such a graphic summary of a pupil's achievement is sometimes called an educational profile.

The graphic record card given on page 57 (Figure 1) represents the record of Robert K., a boy 11 years and 9 months of age on October 12th. At the time the tests were given he was in the 5A grade. This is indicated by the heavy vertical line under the 5th grade. On the mental test which was used he made a score almost equal to a seventh grade child; on reading he scored midway between the sixth and

seventh grade norms. On all but two tests, arithmetic and writing, he exceeded sixth grade norms. Because of his superior mental ability and his excellent work in reading and language, and in spite of his somewhat lower scores in arithmetic and writing it was thought advisable to give him a chance to attempt to carry 6A work. Many other similar uses for this type of graphic record may be worked out by the interested teacher. A very useful individual pupil test record card is given on page 375 of the *Appendix*.

111. Age Scores as Derived Scores—Type 2. One of the most widely used and most meaningful types of derived scores is the age score. This score is obtained by giving the test to all pupils of certain ages which are available for the purpose regardless of the grades in which they may be located. The test scores are then distributed by age groups and the median or average scores computed. The resulting scores for the various age groups are taken as expressing the typical achievement of these groups. For example, if it is found that typical children ten years and three months of age make an average score of 46 points on a certain test, then any other child making a score of 46 points on the test is said to have a test age of 10 years and three months. If the test is a mental test the resulting age score is a mental age. If it is an achievement test the resulting age is an achievement age, the subject age assigned depending upon the subject tested. The age score is fundamental to the development of quotients in the field of measurement.

The use of age norms as a means of giving meaning to test scores is illustrated in Table 38, which is adapted from the *Examiner's Manual* for the *Thorndike-McCall Silent Reading Scales*.[10] This table shows the T-Scores and Reading Ages corresponding to a given number of questions answered correctly by pupils taking the test. The Reading Ages of the three pupils discussed in Table 39 were obtained by

[10] *Thorndike-McCall Silent Reading Scales*, published by Bureau of Publications, Teachers College, Columbia University, New York.

referring to Table 38. This means that Pupil A's score of 19 questions answered correctly is equivalent to a T-Score of

Table 38

TABLE OF *Thorndike–McCall*
T-SCORES AND READING AGES

Number of Questions Right	T Score	Reading Age, Months
36	92	269
35	84	246
34	78	229
33	73	215
32	70	206
31	67	198
30	64	189
29	61	181
28	59	175
27	56	167
26	54	161
25	52	155
24	50	150
23	48	144
22	47	141
21	45	135
20	44	133
19	43	130
18	41	124
17	40	121
16	39	118
15	37	113
14	36	110
13	35	107
12	34	104

43, and a reading age of 130 months. Whether or not this is a good score for Pupil A depends on how this reading age compares with his mental age and chronological age.

The use of age scores in relation to the chronological age of the pupil is presented in Section 113 under the heading of Quotients as Derived Scores. Problem 30 of the "Work-Book in Educational Measurements" illustrates more completely the method of determining age scores from tables of age norms for tests.

Table 39

Pupil	Questions Answered Right	T Score	Reading Age, (Months)
A	19	43	130
B	34	78	229
C	26	54	161

112. Percentile Ranks as Derived Scores—Type 3. The test performance of a pupil may be expressed in terms of the position which his score occupies in the standard distribution of scores for his particular grade. This is accomplished by dividing the distribution so that exact divisions contain the same per cent of the total number of cases. One plan is to divide the distribution into quarters (quartiles), another into tenths, twenty-fifths, hundredths, etc. When the distribution is divided into one hundred parts, each part is called a *percentile.* (See Section 96, Chapter VI.) Ability as represented by a test score is assigned as a position in the rank of one hundred individuals measured by the test. Percentile ranks are particularly valuable where no standards are available, as in the case of most informal objective examinations, or where standards are limited in meaning as is the case with most high school tests. The percentile rank [11] also affords a very convenient means of reducing to a comparable

[11] For a more complete discussion of percentile ranks and the method of computing them see Chapter VI, pages 169 to 173.

basis measures from tests of other school subjects which are otherwise not comparable.

The methods of actually turning test scores into percentile scores are discussed in detail in Section 96 to 99 of Chapter VI. Problems 27 and 28 of the "Work-Book in Educational Measurements" also give additional practical uses of percentile ranks as derived scores.

113. Quotients as Derived Scores—Type 4. Age norms for tests permit the expression of a relationship between the chronological age and the mental and subject ages of pupils. The relationship thus expressed is called a *quotient.*

Intelligence Quotients: When mental age is used as the numerator in the fraction the resulting quotient is called an *Intelligence Quotient.* The following example will illustrate the method of determining the Intelligence Quotient (*I.Q.*). Suppose a pupil exactly eleven years of age earns a score on a mental test which gives a mental age of twelve years and ten months. The *I.Q.* is determined as follows:

$$\frac{\text{Mental Age (in months)}}{\text{Chronological Age (in months)}} = \frac{154 \text{ mos.}}{132 \text{ mos.}} = 1.17 = 117$$

In practice the result of dividing the mental age in months by the chronological age in months is multiplied by 100 to permit the expression of the quotient as a whole number. An *I.Q.* of 100, of course, means perfectly normal development on the part of the individual, within the limits of the reliability of the test, of course. Average *I.Q.*'s range from 90 to 110. *The I.Q. is an index of relative mental development and in no sense expresses absolute mental ability. Its use disregards the level of mental ability, but expresses more exactly the relative brightness of the pupil.* On this account it becomes a valuable instrument for the sectioning of pupils after general classification has been effected by means of mental and educational age data.

Educational Quotients: Many achievement tests permit the determination of quotients in a manner similar to that fol-

lowed in the case of mental tests. Scores from a series or battery of achievement tests such as the *Stanford Achievement Tests,*[12] the *Illinois Examination,*[13] or the *Lippincott-Chapman Classroom Products Test,*[14] for example, afford a very valuable bit of supervisory information when the educational ages and the educational quotients of the various pupils are determined.

The method of turning an educational test score into a quotient representative of the relative development of the individual pupil on the test used may be illustrated by referring to Table 38 on page 199. This table gives the T-Score and Reading Age equivalents for the *Thorndike-McCall Silent Reading Scales.* In connection with the illustration on page 200 the scores of three pupils on this test were used. (Table 39.) The same pupils are utilized here for the purpose of showing how the Reading Quotients are computed. Table 40 summarizes all necessary data on these pupils.

Table 40

Pupil	Questions Answered Right	T Score	Reading Age (Months)	Chronological Age (Months)	Reading Quotient
A	19	43	130	160	81
B	34	78	229	182	126
C	26	54	161	163	99

The score of Pupil A gives him a reading age of 130 months according to Table 38, page 199. He is found to be 160 months of age chronologically. He is therefore not developed in his reading ability so high as might be expected from his

[12] *Stanford Achievement Tests.* Kelley, Ruch, and Terman. World Book Company.

[13] *Illinois Examination,* Monroe, W. S., and Buckingham, B. R., Bureau of Educational Research, University of Illinois, Urbana.

[14] *Lippincott-Chapman Classroom Products Tests.* J. B. Lippincott Co., Philadelphia.

chronological age. His reading quotient is therefore less than normal, or is equal to $\frac{130}{160}$, or 81. The reading quotients for the other pupils in this illustration are found in a similar manner. Problems 31, 32, and 33 of the "Work-Book in Educational Measurements" give additional experience with the calculation of quotients as derived scores.

Accomplishment Quotients: Teachers often raise the issue that a certain pupil is not making the progress in his school work that he is capable of making. When both intelligence and subject tests are given and the scores are turned into mental and subject ages it is possible to answer this question directly. The process is called finding the Accomplishment Ratio or the Achievement Quotient. The educational significance of this technique has been most completely presented and defended by Franzen.[15] The *Accomplishment Quotient* or *A.Q.* is nothing more than the ratio between the relative educational development of a pupil (*E.Q.*) to his relative brightness (*I.Q.*). In symbols this statement takes the form:

$$A.Q. = \frac{\frac{E.A.}{C.A.}}{\frac{M.A.}{C.A.}} = \frac{E.A.}{M.A.}$$ where *E.A.* is the educational age of the subject and *M.A.* is the Binet-mental age of the subject. Since chronological age (*C.A.*) is common to both factors in the ratio the *A.Q.* may be found by dividing *E.A.* directly by *M.A.* If a pupil's achievement (*E.A.*) is in keeping with his power to learn (*M.A.*) his *A.Q.* is 100. Obviously, if the quotient is below 100 the pupil is not working up to capacity. On the other hand a recent investigation[16] of highly motivated instructional drives on certain subject matter tends to show that an *A.Q.* of more than 100

[15] Franzen, R., "The Accomplishment Ratio; a Treatment of the Inherited Disparity in School Product." Teachers College, Columbia University Contributions to Education No. 125; New York, 1922.

[16] Lessenger, W. E., "Motivation and the Accomplishment Quotient Technique." University of Iowa Study in Education, Iowa City, 1925.

is obtainable. It is probable, however, that in the majority of cases, *A.Q.'s* of over 100 will be the result of norms on subject matter tests which are spuriously low. When we have true or correct norms (representative of what is done by normal children under approximately perfect conditions) it seems certain that *A.Q.*'s of more than 100 will disappear. In fact it may not be unlikely that the frequency with which *A.Q.*'s of over 100 are obtained may be used to demonstrate how really far many achievement test norms fail to represent the conditions which norms and standards of achievement imply.

The method of computing the accomplishment quotient may be illustrated if we utilize the records of Pupils A, B, and C, from Tables 39 and 40 on pages 200 and 202. The reading ages of these pupils are available. Let us assume that mental tests have been given and the mental ages recorded in Table 41 are found. This table summarizes all data necessary

Table 41

Pupil	Chronological Age (Months)	Mental Age	*I. Q.*	Reading Age	Accomplishment Quotient
A	160	136	85	130	96
B	182	216	119	229	106
C	163	165	101	161	98

for the calculation of the Reading Accomplishment Quotients for these pupils. The *I.Q.* is the ratio of the mental age and the chronological age of the pupil. The *I.Q.* for Pupil A is $\frac{136}{160}$, or 85. The Reading Accomplishment Quotient is the ratio of the reading age of the pupil to his mental age. For Pupil B this ratio is $\frac{229}{216}$, or 106. In theory an accomplishment quotient of more than 100 is an impossibility, since

it is maintained that a pupil cannot be expected to develop a special skill higher than his general mental ability. The probabilities are that when such quotients exceed 100 the reason for it may be traced to the unreliability of the tests themselves. The Accomplishment Quotient of Pupil C, 98, means that his reading ability as measured by the test used is developed practically up to the level of his mental ability.

Further experience in the finding of quotients from actual school data of the different types is given in Problem 18 on page 374 of the *Appendix* to this volume, and in Problems 30 to 33 in the "Work-Book in Educational Measurements.'

III. Summary

This chapter presents the next logical steps in the interpretation of test results after the necessary statistical summaries have been made. These were discussed in Chapter VI. This chapter brings out definitely the meaning of norms and standards and their use in giving meaning to test scores. The reasons for the need for age norms for tests are concretely illustrated. The practical uses which may be made of so-called derived scores are also presented and illustrated.

The problems involved in the interpretation of the results of testing are many and important. The treatment accorded to them in this volume is so limited by the requirements of space that the attention of the student interested in securing a more critical basis for the interpretation of test results is called particularly to the references given at the end of this chapter.

Chapter VIII takes up specifically the problem of using tests and test results as the basis of remedial and preventive teaching.

SELECTED REFERENCES

Franzen, Raymond, "The Accomplishment Quotient Technique." Teachers College Contributions to Education, No. 125. Columbia University, New York, 1922.

Gilliland, A. R., and Jordan, R. H., "Educational Measurements and the Class Room Teacher." The Century Company, New York, 1925.

Greene, H. A., "A State Survey of Silent Reading in Iowa." Bulletin of Extension Division, Number 91. University of Iowa, 1923.

Gregory, C. A., "Fundamentals of Educational Measurements." D. Appleton and Company, New York, 1922.

Hines, H. C., "A Guide to Educational Measurements." Houghton Mifflin Company, Boston, 1923.

Kelley, T. L.; Ruch, G. M.; Terman, L. M., "The Stanford Achievement Tests." World Book Company, Yonkers-on-Hudson, New York.

Ruch, G. M.; Knight, F. B.; Greene, H. A.; Studebaker, J. W., *The Compass Diagnostic Tests in Arithmetic.* Scott, Foresman and Company, Chicago.

Ruch, G. M., and Stoddard, G. D., "Tests and Measurements in High School Instruction." World Book Company, Yonkers-on-Hudson, New York, 1927. Chapter XIX.

Russell, Charles, "Classroom Tests." Ginn and Company, Boston, 1926.

McCall, W. A., "How to Measure in Education." The Macmillan Company, New York, 1922.

Monroe, W. S., DeVoss, James C., and Kelly, F. J., "Educational Tests and Measurements." Revised and Enlarged Edition. Houghton Mifflin Company, Boston, 1923.

Pressey, S. L., and Pressey, L. C., "Introduction to Use of Standard Tests." World Book Company, Yonkers-on-Hudson, New York, 1922.

Trabue, M. R., "Measuring Results in Education." American Book Company, New York, 1925.

Woody, Clifford, and McCall, W. A., *The Woody-McCall Mixed Fundamentals Tests in Arithmetic,* Class Record Sheets. Teachers College Bureau of Publications, Columbia University, New York City.

CHAPTER VIII

THE PLACE OF TESTS IN DIAGNOSIS AND REMEDIAL TEACHING

The purpose of this chapter is the discussion of the following points which are essential to a clear understanding of the function of tests in the diagnosis of pupil difficulties and the use of this information in remedial teaching:

a. The teacher's responsibility for the diagnostic use of tests.
b. The function of exact diagnosis by means of educational tests.
c. The relation of remedial teaching to exact diagnosis.
d. Principles for the construction of remedial materials.
e. The concrete application of these principles to the construction and use of remedial materials.

I. The Function of Exact Diagnosis

114. Importance of the Diagnostic Interpretation of Test Results. There is nothing mysterious in the use of standard or informal objective tests which results in the immediate improvement of instruction in the classroom. As has been repeatedly recognized and stated by workers in this field, tests as such are incapable of improving instruction because of any inherent power of their own. Existing conditions are merely revealed by them, and these with the limitations implied by the validity and reliability of the particular instruments used. Remedial or corrective teaching is the result of deliberate constructive effort by the teacher after the particular points of weakness in the instruction of the pupils

have been revealed by the tests. The ease, clearness, and directness with which these needs are revealed by the test is a measure of its real educational value. Too few of existing standard tests are so constructed as to permit the interpretation of their results directly in terms of an effective remedial procedure. However, this seems to be no good reason for the failure of teachers to apply more directly the results of this work in testing to the improvement of their teaching practice. Just as the data revealed by the navigator's instruments require calculation and interpretation, so is it necessary to analyze carefully test data in order to make it the basis of a genuine remedial program. Nor is the teacher the only one to profit from a frank study of test scores. Pupils themselves will often take an active part in remedial enterprises when they are given the facts concerning their own performances on standard tests.

The interpretation of test scores and the planning of remedial procedures is the most difficult part of the use of standard educational tests. Moreover, it is by far the most important part. One of the greatest needs in education today is the provision for genuine diagnostic testing in all instructional fields, supplemented by valid remedial work designed to correct the weaknesses and defects of individual pupils as revealed by the tests. Up to the present time the makers and users of tests have been for the most part content merely to measure. It is coming to be recognized that the giving of a test with a study of the resulting scores will not in itself improve the classroom situation. This is especially true if the test is of the general survey type. It is interesting to learn as a result of using tests in the classroom that an individual pupil or the entire class is below standard in the subject, but unless we learn with some exactness what causes the low level of achievement our testing program will do little if anything more than supply us with interesting information. Teachers and supervisors have a right to expect that something more constructive will be provided in exchange for the time required for classroom testing.

115. Exact Diagnosis the Basis for Remedial Work. Exact diagnosis of class and individual pupil difficulties coupled with the application of specific remedy is the heart of enlightened use of exact methods of teaching. It is the logical outgrowth of the measurement movement in education. The success of the remedial or corrective teaching depends upon the accuracy and detail with which the specific skills involved in the successful achievement in the subject are identified and isolated in the test. General survey tests can not supply this information in sufficient detail. This is illustrated by the record of a single pupil given in Table 42.

Table 42

TEST RECORDS OF PUPIL A. L. C.
(Grade 4A, School L)

	Standard or Norm	Pupil Score
Courtis Series B, Addition:		
Attempts	6	3
Rights	6	1
Accuracy	100%	33%
Compass Diagnostic Test in Arithmetic:		
Test 1. Addition of Whole Numbers:		
Part 1. Basic Addition Facts	62	60
Part 2. Higher Decade Addition	60	58
Part 3. Column Addition	40	24
Part 4. Carrying in Column Addition	80	30
Part 5. Checking Answers in Addition	31	20
Total Score	273	192

This table shows the test scores of Pupil A. L. C., a fourth grade pupil in School L, on the *Courtis Series B Arithmetic Test, Addition,* and on the *Compass Diagnostic Tests in Arithmetic, Test 1, Addition.* His score of 3 examples at-

tempted in the Courtis Test and one worked correctly indicates that he is having some type of difficulty with his work in addition. The fact that he is able to work only 3 of these examples as compared with 6, the standard for his grade is significant but not particularly enlightening to his teacher, who accordingly turns to a more detailed measuring instrument and gives him also the *Compass Diagnostic Test Number 1, Addition of Whole Numbers.* His scores on this test are enlightening, as is shown in Table 42. Of the 70 basic addition facts listed in Part I of this test he is able to give the correct answers to 60 in the time allotted. Apparently his difficulty is not at this point. Of the higher decade addition facts tested in Part 2 his score of 58 is also satisfactory for his grade. However in Part 3, Column Addition, his difficulty begins to appear, as shown by his score of 24 points compared with the grade norm of 40. He seems to know the higher decade facts in isolation but when he is called upon to carry these unseen sums in his mind his attention breaks down, and he forgets his result and accordingly is obliged to begin all over again. This makes for slow and inaccurate work in column addition of the type comprising the Addition Test of the *Courtis Series B Test,* and Parts 4 and 5 of the *Compass Diagnostic Test, Test 1,* a situation which is not revealed at all by the survey type of test.

It must be clear, therefore, that highly detailed diagnostic instruments must be provided, capable of sampling into the underlying skills in a special field adequately to furnish a reliable individual diagnosis. Clearly the development of such material implies an accuracy and detail of analysis of the skills underlying the different school subjects which is far in excess of present accomplishment in the majority of these fields. Skills which can not be identified can scarcely be improved except by mere good luck or by "blunderbuss" methods of instruction. Frequently, however, such methods are effective, due to the fact that a large enough charge of drill fired in the general direction of a pupil's difficulty may strike the weakness by pure chance. Remedial teaching

can be made effective only when the exact point of weakness causing the breakdown in pupil mastery has been located. On this account the analysis of the basic skills must be detailed, and the diagnosis must provide accurate, detailed, and reliable information concerning these skills.

In connection with their "Diagnostic Chart for Individual Difficulties in the Fundamental Processes in Arithmetic," Buswell and John [1] have prepared a very interesting and valuable list of habits of work followed by children in their arithmetic. The type of diagnosis proposed by Buswell and John is an individual type. The teacher or examiner watches the pupil as he works on a specially arranged individual test sheet, and records observations and notes on the pupil's habits of work as he solves the example. While in a sense these items (habits) represent the negative aspects of skills in arithmetic they nevertheless offer some interesting suggestions as to the desirable positive skills. A number of the work-habits of pupils engaged in solving ordinary addition examples are reproduced on page 238 in Chapter IX as an illustration of the exact type of analysis which must precede diagnostic and remedial treatment in arithmetic and other similar fields.

116. Diagnosis as the Basis for Preventive Work. An examination of the number and types of skills identified as a result of the diagnostic methods discussed above leads to the suggestion of a still further use of diagnostic test results. The function of diagnosis is to locate specific difficulties and if possible to determine their causes. Out of this should come the basis for the development of instructional methods having as their purpose the prevention of weaknesses through the anticipation of causes. If a parallel may be drawn from the field of medicine the case may be made somewhat more concrete. In every medical examination for diagnostic pur-

[1] Buswell, G. T., and John, Lenore, "Diagnostic Chart for Individual Difficulties; Fundamental Processes in Arithmetic." Public School Publishing Company, Bloomington, Illinois, 1925.

poses a complete case history is kept. Out of the analysis of these records has come a better understanding of the causes and characteristics of certain types of human ailments. Out of this same analysis has also come the basis for much of the marvelous preventive work which characterizes modern medical science. In a similar way accurate and detailed educational diagnosis offers the basis for the development of a program of preventive work in education.

117. General Practice Exercises vs. Remedial Drill. There are in general two ways of maintaining a high level of pupil achievement in any subject after direct instruction has been discontinued. These are (1) systematic drill in general with no integral units of testing to discover breakdowns in pupil mastery, and (2) systematic remedial drill devices to fight forgetting, plus diagnostic testing to discover the exact causes of weaknesses when such weaknesses begin to cause poor work on review drills. The first method involves the systematic use of properly distributed general practice over the complete function. The second involves the periodical location of the specific defects of individual pupils by means of diagnostic tests and the immediate correction of these defects by the use of properly constructed remedial drill.

Unquestionably the latter is the more economical method of maintaining mastery of desired skills on the part of a pupil. It is obvious that general review is valuable at times but just to review with no specific idea of what the review is to accomplish is too naive and hopeful to be effective. The program which coincides most closely with the experience of successful teachers and with a sound psychology of learning calls for the following steps in approximately the order indicated: (1) *teach,* (2) *review,* (3) *test for weaknesses* whenever they appear, and (4) *follow with remedial drill units* on the specific weaknesses revealed by the tests.

Remedial material, since it is quite highly specialized, must be prepared in keeping with certain definite specifications. The criteria for remedial drill listed below are presented as

a general guide to the teacher and supervisor in the preparation or selection of such material. They are believed to be sufficiently general to afford valuable guidance in practically all of the habit forming fields of elementary education.

II. Criteria for Construction of Remedial Material

118. Analysis must Precede Diagnosis.[2] The successful development of the many sets of habits which constitute the bulk of our elementary school teaching depends upon the care with which the underlying and basic habits themselves are recognized and utilized in the teaching. If, in arithmetic, a habit of responding correctly to the 100 basic addition facts is essential to success in addition then this fact must be recognized and stress laid on the development of these skills. If, in addition to these basic skills it can be shown that there are other related but somewhat higher skills which are fundamental also to successful work these also must be recognized and developed. If, for example, teaching a child to add can be shown to consist not only in the mastery of the 100 basic combinations but can also be shown to involve higher levels of skill such as knowledge of the higher decade addition facts, the bridging of the tens, the control of the attention span, carrying from one column to the next, the teacher's task is made much more obvious. Blind trust in general practice on the total skill will give away to an earnest effort to discover the particular point of weakness for special emphasis.

Good diagnosis and the development of valid remedial materials must parallel exactly the processes of good teaching. Effective remedial materials in any school subject can not be prepared until the skills contributing to success in this field have been isolated and identified. Psychologically the reason for this is that on the whole the child learns to do what he practices and not something else. Accordingly, reme-

2 Greene, H. A., "Criteria for Remedial Exercises in Arithmetic," Fourth Yearbook of the Department of Superintendence, pp. 213–219.

dial work, if it is to function, can do so only when the exact level at which pupil mastery breaks down has been located. Thus the analysis must be penetrating and the diagnosis must be exact.

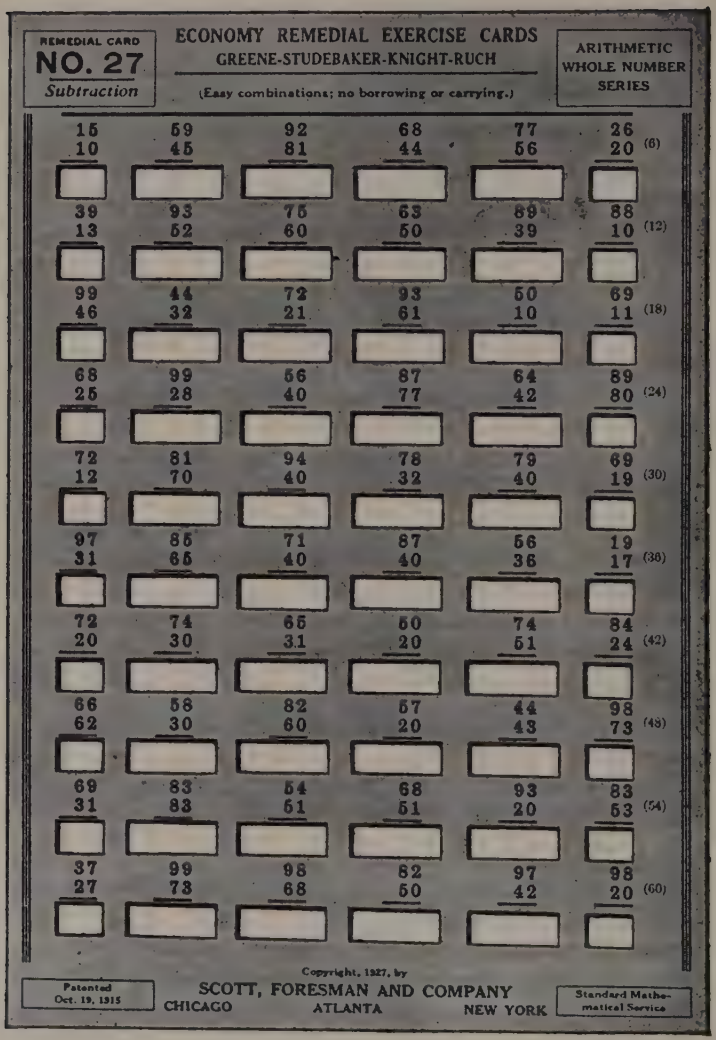
119. Diagnosis Must Be Specific. Diagnosis must be more exact than broad statements of general functions. It is not enough to discover that a child is unable to read silently. The exact nature of his handicap must be revealed before it is possible to undertake a remedial program.[3] The more specific the diagnostic information revealed the more exactly the remedial material can be made to fit the need. To return to a frequently used illustration, it may be shown that the child is unable to add, but unless the exact point at which his mastery of addition breaks down can be determined by the diagnosis, teaching or remedial efforts are largely wasted. One of the outstanding reasons why more effective teaching and remedial work has not been done in certain fields is that no adequate analysis of basic skills has been made. Concrete illustrations of this need are given in connection with a related discussion in Section 120 of this chapter.

120. Each Identified Skill Must Be Paralleled by Valid Drill. Drill material of known validity must be provided for each specific skill which conditions achievement in a specific subject if remedial work is to be effective. The validity of drill material depends to a large degree upon the accuracy and skill with which the analysis of skills is made. Difficulties in subject-matter units which can be only hazily identified can not be remedied except by mere good luck. Drills must closely parallel the skills which they are supposed to remedy. If mastery of a certain minimal vocabulary is essential to effective silent reading comprehension then drill on those particular words which constitute special weakness should preclude all other drill.

[3] The one exception to this is when the child is weak in every skill constituting reading ability. In this case he does not need remedial drill on certain skills, but complete re-teaching of the total process.

Perfect validity of drill material can be achieved only by taking a 100% sampling of all of the possible basic facts or skills in the particular field. This is impossible in many cases, but it is nevertheless often possible to take such a large sampling that all of the most frequently used facts are included. An illustration of each of these cases may be taken from certain remedial units already in existence. In a certain remedial unit in arithmetic which may be called Unit Number 27, there are 55 possible combinations by which the numbers 0 to 9 may be subtracted without involving borrowing or carrying. In this remedial unit each possible combination is sampled twice. Table 43 on page 217 shows how the practice is distributed in the illustration in such a substantially adequate sampling.

This distribution of practice means that in the remedial unit presented above each combination appears twice. Zero is taken twice from one, two, three, etc., and all remaining digits. Moreover, this particular unit is devised in such a way that the child has practice on simple subtraction problems involving a wide range of differences (answers). Even the frequency with which the different integers themselves are used in the differences is under control. Remedial material of this character is more effective than merely random exercises, assuming that proper motivation for improvement is provided. It is almost 100% valid and covers a set of skills of basic importance in the hierarchy of habits upon which successful achievement in subtraction depends. This is, of course, only a small part of total subtraction ability but if the weakness lies here the remedial work should be given here and not at some other point in the instruction. This represents one of the differences between good and poor remedial material. Properly organized and administered remedial drill material strikes directly at the heart of the trouble and wastes no time on skills which need no practice. Poorly organized remedial drill may or may not deal with all of the possible weaknesses, but is almost certain to waste time on skills which are not in need of drill.



Specimen Remedial Card in Subtraction. Reproduced by permission of the publishers.

Table 43[4]

Remedial Unit Number 27
Easy Combinations; No Borrowing Nor Carrying

Remedial Unit 27
Minuend

Subtrahend	0	1	2	3	4	5	6	7	8	9
0	2	2	2	2	2	2	2	2	2	2
1	*	2	2	2	2	2	2	2	2	2
2	*	*	2	2	2	2	2	2	2	2
3	*	*	*	2	2	2	2	2	2	2
4	*	*	*	*	2	2	2	2	2	2
5	*	*	*	*	*	2	2	2	2	2
6	*	*	*	*	*	*	2	2	2	2
7	*	*	*	*	*	*	*	2	2	2
8	*	*	*	*	*	*	*	*	2	2
9	*	*	*	*	*	*	*	*	*	2

Distribution of Differences	Differences by Integers
N	
0 – 5	0 – 18
5 – 3	1 – 20
10 – 7	2 – 16
15 – 3	3 – 13
20 – 7	4 – 12
25 – 3	5 – 11
30 – 5	6 – 8
35 – 3	7 – 6
40 – 4	8 – 4
45 – 2	9 – 2
50 – 6	
55 – 1	110
60 – 2	
65 – 1	
70 – 2	
75 – 1	
55	

[4] Table 43 is interpreted as follows: The subtraction combination 0–0 appears twice in this remedial unit; 1–0 appears twice; 2–1 appears twice, etc., each simple subtraction fact in which borrowing or carrying is not involved is sampled twice. The differences or answers are also distributed in such ways as to afford a wide range of experience. For example, answers ranging from zero to four inclusive are found 5 times; answers ranging from 5 to 9 inclusive are found 3 times, etc. The column headed Differences by Integers indicates the number of times each integer from 0 to 9 appears in the differences (answers). Naturally the larger amount of emphasis must be given to the smaller numbers since there are more opportunities for these smaller differences to arise.

An illustration of the case in which a complete and perfect sampling of material is not possible may be taken from the field of problem solving. A field as complicated as this and as highly related to general intelligence offers many difficulties in the way of a complete analysis and identification of the underlying or basic skills. In general at least five fundamental steps in problem solving closely paralleling the steps in the thinking process as set forth by Dewey [5] are noted. The first step in the solution of the verbal problem involves a complete understanding of the items, elements, and processes which are involved or implied in it. This is *comprehension.* This in itself involves many factors such as rate of reading, vocabulary difficulties, reading of numerals, problem organization as well as complexity in terms of the number and order of the arithmetical processes involved. Underlying all of these is, of course, the ability of the child to hold the various facts and conditions in his mind long enough to analyze and organize them. This process of *analysis and organization* constitutes a second important step. The unnecessary facts or implications are discarded and only the data which really affect the problem are retained. The third step in practice is actually a part of the second, for the *recognition of the process* involved is really a part of analysis. From this the subject moves straight to the fourth step, *solution,* where he applies to a specific situation his knowledge of the fundamental tools of number. In his earlier practice he has learned how to perform certain simple arithmetical skills. Now he learns when to apply them. The next and final step in the process is *verification,* which may be either a rough checking by the estimation of the probable answer to the problem, or by an actual recalculation and re-check of the processes involved.

Of course, many children fail to proceed in the solution of problems in as orderly a fashion as indicated in this dis-

[5] Dewey, John, "How We Think." D. C. Heath and Company, Boston, 1910.

cussion, not because it would not be economical to do so but because they do not see the most direct methods to follow. Oftentimes, imperfect work (though finally successful) means using unnecessary steps, and spending useless energy in doing essential steps in an ineffective order.

So far in the development of diagnostic instruments for the identification of the skills involved in problem solving only a rough sampling of these skills has been approximated. The only two attempts of any consequence in this field are those represented by *Stevenson's Problem Analysis Tests* and by the authors of the *Compass Diagnostic Tests.* Of these, the latter represents the most comprehensive and reliable analysis. The Stevenson test presents only six problems and by means of exercises secures response to four types of questions:

A. Which of the following facts are given in the problem?
B. Which of the following facts are you asked to find out in the problem?
C. Which of the following is the most reasonable answer?
D. Which process should be used in solving the problem?

The *Compass Tests* (Test 17 and 18) present 15 problems on which exercises covering (1) Comprehension, (2) What is Given, (3) What is Called for, (4) Probable Answer, and (5) Correction Solution, are given. In both of the tests an attempt has been made to parallel as closely as possible the above introspective analysis of the skills involved but only for purposes of identification. The arrangement of the exercises in the *Compass Diagnostic Tests on Problem Solving* (Test 17,[6] for grades 5 and 6) is illustrated in the accompanying reproduction of Problem 1.

[6] In the test itself the exercises are arranged so that the pupil works across the page to the right, answering all of the exercises dealing with a given problem before taking up another problem. In this reproduction the size of the page makes this type of arrangement impossible.

Specimen Exercises from Compass Diagnostic Test, Number 17

PROBLEM ANALYSIS

Problems	Part 1—Comprehension	Part 2—What Is Given
Read each problem below. Then work across the two facing pages to the right, doing all the Parts for one problem before going to the next. Do not go back and work on a Part after you have completed the one following. *Read the Sample below.*	*Put a cross (×) on the line before the* **one** *statement below which is true for each problem.*	*Put a cross (×) on the line before* ***every*** *statement below which tells a fact given in the problem.*
Remember: Work across the page to the right.		
[*Read the problem*]	[*Check true statement*]	[*Check what is given*]
Problem 1 *Our baseball team played 7 games this summer. We lost 2 and tied none. How many games did we win?* ——→	—— Team won all games played. —— Team lost all games played. —— Team won more than it lost. —— Team lost half of the games played. —— Team won about half of the games played. ——→	—— Number of games played. —— Number of boys on team. —— Number of games tied. —— Number of games won. —— Number of games lost. ——→ (*From here go to Part 3 below*)

Part 3—What Is Called for	Part 4—Probable Answer	Part 5—Correct Solution
Put a cross (×) on the line before the ***one*** *statement below which tells what is called for in the problem.*	*Put a cross (×) on the line before the* ***one*** *statement below which gives the nearest probable answer to the problem. Do not take time to work the problems.*	*Put a cross (×) on the line before the* ***one*** *correct solution given for each problem. Figure in the margin if you want to.*
Remember: Work across the page to the right.		
[*Check what is called for*]	[*Check probable answer*]	[*Check correct solution*]
—— Number of games lost. —— Number of boys on team. —— Number of games won. —— Number of games where score was tied. —— Number of games played. →	—— 9 boys. —— 6 schools. —— About 7 boys. —— 9 games. —— About 5 games. →	—— $7+2=9$ —— $7-2=5$ —— $7\div3=3\frac{1}{2}$ —— $7\times2=14$ —— $7\times2=14;\ 14-9=5$ Now Start Problem 2.

Reproduced with permission of publishers, Scott, Foresman and Company, Chicago. Copyright, 1925, by publishers.

The remedial aspect of the field of problem solving largely remains to be developed although some definite constructive work on the preparation of instructional and remedial material has been done. One of the more extensive efforts to provide the teacher of arithmetic with practice exercises in problem solving is seen in the "Economy Problem-Solving Exercises" [7] which are designed to parallel very closely the organization of skills measured by the *Compass Diagnostic Tests in Arithmetic,* Tests 17, 18, 19, and 20. Since a complete sampling of the complex mass of skills involved in problem solving can not be made, only a few of the most important ones are included in the proposed remedial material. Practice on the silent reading comprehension of verbal problems of varying degrees of complexity is given. Practice on the selection of the items given in the problem essential to its solution is afforded by these exercises. Skill in comprehending the real problem setting by determining what is called for in the problem is also developed by much practice on this type of exercises. Practice on the basic skill of choosing the correct process or combination of processes in the more complex problems is also given. Skill in the verification of the solution or the estimation of the most probable answer to the problem is developed by a special set of exercises. The complete set of problem exercises can be utilized as a means of unifying the skills involved in the complete process of problem solving. Table 44 on page 223 shows in compact form the relationship between the steps involved in problem solving, the factors underlying success in problem solving, and the types of drill provided by the "Economy Problem Solving Exercises." [8]

In this manner at least six of the basic skills of problem solving are provided with remedial drill. The validity of the drill depends upon the degree to which this sampling

[7] "Economy Problem-Solving Exercises," by Greene, H. A., Studebaker, J. W., Knight, F. B., and Ruch, G. M. Scott, Foresman and Company.

[8] *Ibid.*

Table 44

Analysis of Problem Solving

Steps in Problem Solving	Factors Underlying Problem Solving	Types of Drill Provided
Comprehension	Vocabulary Ability to Read Numerals Ability to Read Rapidly Ability to Comprehend *a.* Follow directions *b.* Make generalizations *c.* Select potent elements *d.* Discard irrelevancies *e.* Determine problem setting as a unit *f.* Determine the outcome of the problem *g.* Grasp the significance of problem cues	Multiple Choice Comprehension Exercises
Analysis and Organization	Selection of Potent Elements Selection of Processes Involved Determining What the Problem Calls for Determining What is Given in the Problem Determining the Process Relationships	What is Given [9] Process Analysis What is Called for Problem Relationships
Recognition	Choice of Procedure Determining Problem Conditions Determining the Purpose of the Problems Determining Relevant Elements	Process Analysis What is Called for What is Given
Solution	Selection of Process Organization of Processes in Order Knowledge and Application of Combinations Problem Relationships	Process Analysis Problem Scales
Verification	Probable Form of Answer Probable Magnitude of Answer	Probable Answer

[9] The order of presentation of these types of drill bears no relation to the order of the factors underlying problem solving.

covers the basic or fundamental skills and the degree to which the exercises themselves actually develop the skills they purport to develop. There are a number of places in which this complex chain may break. The task of diagnostic and remedial treatment is to locate and repair quickly those links of the chain which have snapped under stress, or have rusted out through lack of use.

121. The Complete Area of the Skill Must Be Covered. Correctly designed remedial material will not only parallel with valid drill the correct skills but it will also cover all of the basic aspects of the skill. Remedial drill in subtraction of fractions will cover as far as possible all types of situations which arise. Furthermore it should acquaint the child with the most important variants of each situation. For example, in the development of skill in the subtraction of fractions, provision must be made for drill on the finding of common denominators, reduction of proper fractions, the reduction of improper fractions, the borrowing (or carrying) difficulty, etc.

Specimen Copy of Fractions Remedial Unit

Choosing Common Denominators

Directions.—You can always get a common denominator by multiplying the denominators together. But sometimes smaller denominators can be used. Always use a small common denominator when you can. Practice on this card until you can choose denominators very quickly and accurately.

Write the denominator you would use in each example below:

Sample: $\frac{}{2}+\frac{}{4}+\frac{}{3}=\frac{}{12}$

$\frac{}{2}+\frac{}{16}+\frac{}{8}=$	$\frac{}{12}+\frac{}{4}+\frac{}{3}=$	$\frac{}{6}+\frac{}{15}+\frac{}{5}=$	(3)
$\frac{}{4}+\frac{}{6}+\frac{}{2}=$	$\frac{}{6}+\frac{}{6}+\frac{}{9}=$	$\frac{}{5}+\frac{}{3}+\frac{}{3}=$	(6)

$\frac{}{8}+\frac{}{12}+\frac{}{4}=$	$\frac{}{2}+\frac{}{4}+\frac{}{12}=$	$\frac{}{18}+\frac{}{3}+\frac{}{6}=$	(9)
$\frac{}{10}+\frac{}{4}+\frac{}{5}=$	$\frac{}{5}+\frac{}{4}+\frac{}{2}=$	$\frac{}{6}+\frac{}{12}+\frac{}{8}=$	(12)
$\frac{}{3}+\frac{}{2}+\frac{}{12}=$	$\frac{}{2}+\frac{}{16}+\frac{}{4}=$	$\frac{}{8}+\frac{}{3}+\frac{}{4}=$	(15)
$\frac{}{24}+\frac{}{2}+\frac{}{8}=$	$\frac{}{6}+\frac{}{5}+\frac{}{10}=$	$\frac{}{12}+\frac{}{6}+\frac{}{9}=$	(18)
$\frac{}{2}+\frac{}{5}+\frac{}{3}=$	$\frac{}{8}+\frac{}{4}+\frac{}{2}=$	$\frac{}{2}+\frac{}{16}+\frac{}{8}=$	(21)
$\frac{}{5}+\frac{}{2}+\frac{}{5}=$	$\frac{}{12}+\frac{}{8}+\frac{}{4}=$	$\frac{}{6}+\frac{}{3}+\frac{}{16}=$	(24)

An illustration of an exercise giving practice on the finding of common denominators in the subtraction of fractions is given herewith. This is one of a series of such exercise cards projected for use in the correction of difficulties in the teaching of fractions.

The need for special care in the preparation of such material can be shown clearly by referring to Table 45 on page 227 which presents a distribution of the practice on a special remedial unit dealing with the subtraction of whole numbers. This particular unit is so constructed that the child is confronted with sixty operations in subtraction involving borrowing, forty-five of which are apparent to the child and fifteen of which do not appear until after the first borrowing has taken place. Not only does it provide drill on all of the possible combinations but it places special emphasis on the more difficult ones. In addition it gives the child experience with differences of various magnitudes as well as experience in borrowing from integers of different magnitude. In other words a special effort has been made to cover the entire area of the borrowing difficulties involved in subtraction of whole numbers. Equal care must be exercised in the development of similar material for other fields.

REMEDIAL CARD
NO. 31
Subtraction

ECONOMY REMEDIAL EXERCISE CARDS
GREENE-STUDEBAKER-KNIGHT-RUCH
(Varied types of borrowing or carrying.)

ARITHMETIC
WHOLE NUMBER
SERIES

211 109	310 209	493 337	322 109	284 117	(5)
270 147	459 191	519 240	340 116	395 217	(10)
456 166	687 293	718 383	391 257	282 137	(15)
496 307	527 181	709 352	410 208	649 281	(20)
648 164	361 126	392 146	829 450	764 189	(25)
713 146	483 226	776 281	668 182	883 297	(30)
854 295	944 157	737 290	812 118	1177 278	(35)
1164 396	638 180	814 338	877 299	1345 447	(40)
1266 389	921 129	938 249	819 350	1575 586	(45)

Copyright, 1927, by

Patented
Oct. 19, 1915

SCOTT, FORESMAN AND COMPANY
CHICAGO ATLANTA NEW YORK

Standard Mathematical Service

Specimen Remedial Card in Subtraction. Reproduced by permission of the publishers.

Table 45

REMEDIAL UNIT NUMBER 31
More than One Borrowing or Carrying

Distribution of Practice

Subtrahend \ Minuend	10	11	12	13	14	15	16	17	18
0	*	*	*	*	*	*	*	*	*
1	1	*	*	*	*	*	*	*	*
2	1	1	*	*	*	*	*	*	*
3	2	1	*	*	*	*	*	*	*
4	1	1	2	1	*	*	*	*	*
5	1	1	1	1	2	*	*	*	*
6	1	1	1	2	2	2	*	*	*
7	1	1	1	2	2	2	2	*	*
8	1	1	2	1	2	2	2	2	*
9	1	2	1	1	2	2	2	2	

Distribution of Differences	Cases of Borrowing From
N	*N*
50— 1	0— 0
100— 4	1— 6
150— 3	2— 2
200— 6	3— 1
250— 4	4— 5
300— 3	5— 3
350— 3	6— 9
400— 1	7—10
450— 6	8—11
500— 0	9— 8
550— 5	10— 0
600— 0	11— 2
650— 2	12— 1
700— 0	13— 1
750— 3	14— 0
800— 0	15— 1
850— 3	—
900— 0	Total 60
950— 1	
—	Sixty operations in borrowing 45 seen 15 unseen
Total 45	

122. The Underlying Skills must be Synthesized in Remedial Units. Effective remedial material must not only cover in a valid manner all of the basic or underlying skills upon which achievement in the field depends but it must provide a means for bringing about a gradual union of these component elements into the total function. It is entirely possible that a mastery of the subsidiary skills involved might result in only a partial mastery of the end-product if that goal were

not reached by the gradual bringing together of each distinct skill into its relation to the whole skill. This point may be illustrated by referring to the following tentative list of specific skills involved in mastery of the long division process. Experimentation and observation show that a breakdown in the total skill may take place at any one of the following points:

1. Knowledge of vocabulary of long division
2. The form of long division
3. The fundamental division combinations
4. Carrying in long division
5. The multiplication involved in long division
6. The addition involved in long division
7. The subtraction involved in long division
8. The estimation of quotient figures both apparent and non-apparent.[10]

From this it appears that there are at least eight basic skills and concepts which must be developed and brought into relation with each other if mastery of long division is to be attained. However, in order to provide adequately for the synthesis of these skills into the complete process at least ten more remedial units are necessary.[11]

123. Time and Achievement Standards Should Accompany Remedial Drill Material. The establishment of time and

[10] Quotients in long division are said to be *apparent* when the trial quotient resulting from dividing the left or first digit of the divisor into the first figure (first and second if the divisor will not go into the first figure) of the dividend is the actual or true quotient, as in $12\overline{)372}^{\,13}$. Here 1 goes into 3 exactly three times and the true first quotient figure is 3. Quotients are said to be *non-apparent* when the trial quotient obtained in a similar way is not the true quotient, as in $17\overline{)323}^{\,19}$. Here the 1 goes into 3 three times, but the true first-quotient figure is 1 and not 3 as it appears to be on first trial. These types of estimation of quotient figures account for a large part of the difficulty pupils encounter in long division.

[11] Greene, H. A., Studebaker, J. W., Knight, F. B., and Ruch, G. M., "Economy Remedial Exercises in Whole Numbers, Teachers Manual." Scott, Foresman and Company.

achievement norms for remedial material enables the child to discover for himself when he has reached a reasonable degree of proficiency as a result of drill on a specific skill. It also affords a valuable motivating agency for the child and permits the evaluation of improvement in the skills in terms of units of time. The teacher will, of course, be interested especially in this latter aspect.

No particular significance is to be attached to the fact that many of the illustrations in the foregoing discussion of criteria for remedial materials have been taken from the field of arithmetic. The field of arithmetic happens to be a field in which rather complete identification of the basic skills has already been made. When other fields have been as carefully analyzed the criteria will apply as clearly there as in the field of arithmetic.

III. Summary

This chapter is in a very real sense the point at which all of the foregoing discussion of the meaning and uses of tests as instruments for the improvement of classroom instruction has been aimed. The chapter should clinch in the mind of the teacher and student the often repeated fact that the underlying purpose of all testing is the accurate determination of class and individual pupil difficulties to the end that remedial instruction may follow. It is shown by means of concrete illustrations that exact and accurate diagnosis of specific pupil weaknesses affords the only dependable basis for effective remedial work. General principles for guidance in the construction of remedial materials are presented and illustrated.

The treatment of diagnostic testing and remedial teaching given in this chapter is purposely presented from a general point of view. The specific application of diagnostic methods and the development of remedial materials in certain subject matter fields is discussed in succeeding chapters. However, materials for exact diagnosis and effective remedial

instruction have not been equally well developed in all fields. Since the purpose of this book is to show in a simple and concrete way how tests may be used as an instructional aid to the classroom teacher the actual application of diagnostic methods and remedial instruction is made in the fields in which the greatest amount of material is available, namely, in the fields of arithmetic, language, and reading. Similar applications of diagnosis and remedial procedures may be made in other fields as rapidly as the subjects themselves are submitted to the detailed analysis necessary to identify the underlying skills and sufficiently detailed and reliable diagnostic instruments and corrective materials are developed.

Chapter IX takes up the problems of remedial teaching following diagnosis as applied to the field of arithmetic in particular.

SELECTED REFERENCES

Dewey, John, "How We Think." D. C. Heath, Boston, 1910.

Greene, H. A., "Criteria for Remedial Drill in Arithmetic." Fourth Yearbook of the Department of Superintendence, N. E. A., February, 1926.

Greene, H. A.; Studebaker, J. W.; Knight, F. B.; Ruch, G. M., "Economy Problem Solving Exercises." Scott, Foresman and Company, Chicago, 1928.

Greene, H. A.; Studebaker, J. W.; Knight, F. B.; Ruch, G. M., "Economy Remedial Exercises in Arithmetic, Manual for Whole Number Series." Scott, Foresman and Company, Chicago, 1927.

CHAPTER IX

DIAGNOSTIC AND REMEDIAL TECHNIQUES; ARITHMETIC

The purpose of this chapter is to summarize the following points involved in the diagnosis and correction of pupil difficulties in the field of arithmetic:

a. The possible scope of remedial work in arithmetic.
b. The teacher's responsibility for the use of remedial methods.
c. Specific skills in arithmetic which may be diagnosed and remedied.
d. Types of remedial materials in arithmetic.
e. The classroom use of remedial exercises.

124.—Introduction. It has been repeatedly pointed out in preceding chapters of this book that tests as such are incapable of improving instruction. Existing conditions are merely revealed by them, and it is worthy of note that these conditions are revealed only within the limits of the validity and the reliability of the particular tests used. Repeated emphasis has been placed throughout the pages of this volume on the importance of using measuring devices which are themselves based upon a sufficiently detailed analysis of the basic skills required for successful achievement in the field to permit the application of definite remedial procedures. Remedial teaching is the result of deliberate instructional effort on the part of the teacher after the particular points of weakness of the pupils have been revealed by tests of one kind or another. The accuracy with which these needs are revealed by

the instrument used is the best measure of its value to the classroom teacher. Specific criteria for the evaluation of test materials as such are discussed in Chapter IV. Suggested criteria for the evaluation of available material designed for remedial instruction purposes are given in Chapter VIII. The same principles are certain to be useful to the student or teacher who is interested in the preparation of her own remedial materials.

125. Scope of Remedial Teaching in Arithmetic. The criteria for remedial exercises set forth in the preceding chapter, while representing the principles on which the construction of remedial exercises should be based, at the same time make clear certain factors which have tended to interfere with the more general development of such material. It is not just chance that diagnostic tests have been developed in subject matter fields where the aims are clean cut and the basic skills upon which achievement depends have been analyzed carefully. Nor is it chance that the blanket purposes of certain other subject matter fields as expressed in courses of study and textbooks have left the teacher groping vaguely for tangible goals and effective instructional methods. The order of development is clear; first, there must be a specific statement of aims lying back of the subject matter; second, a detailed analysis must be made of the basic skills upon which ultimate achievement depends; and third, there is the preparation of material designed to give mastery of these skills.

Some progress has been made in the diagnosis and remedy of pupil defects in the field of arithmetic. This is largely because the aims of arithmetic are quite clearly stated, which in turn makes possible a rather detailed analysis of the underlying skills. As long as a school subject has for its major purpose the accomplishment of such indefinite aims as the development of the "cultural" and "disciplinary" values of life nothing definite in the way of discovering existing difficulties in instruction or of undertaking ways of re-

moving these difficulties can be done. However, as soon as it becomes known, as it is now known, for example, that the ability to do a certain type of column addition is a desirable outcome in arithmetic and depends upon the pupil's knowledge of approximately 450 higher decade addition facts, it becomes possible not only to locate difficulties in teaching this material as such, but it also is possible to furnish the teacher with specific aids in teaching it. Formerly it was assumed that because a child happened to know that 9+8=17, he would naturally know that 39+8=47. This we now know to be an unsafe assumption. Thus, such materials as the 450 higher decade addition facts constitute valuable teaching and remedial material. The reason why similar material is not available in such fields as geography, history, science, etc., is that the aims of instruction in these fields have not yet become sufficiently crystallized to permit the type of analysis which has been given arithmetic.[1]

I. Identification of Arithmetical Skills

126. Specific Skills in Arithmetic as Basis for Test and Remedial Units. In connection with the discussion of the problems of diagnosis in their relation to remedial work presented in Chapter VIII it was pointed out that the problems and the materials needed for the first teaching of a subject and for remedial instruction in the same field are closely related. In general, the methods and the materials which are effective in the first teaching of a skill are useful in the

[1] It should probably be pointed out here that it is likely that there are certain subject matter fields in which this type of crystallization of aims will not and should not take place. This is undoubtedly true of certain subjects such as social science and the natural and physical sciences in which changes in content are appearing so rapidly and in which there are certain points on which general agreement can not be expected. Here the diagnosis will remain for some time in general terms, such as the ability to read or the ability to work the mathematics involved in certain science fields. Here also remedy will be largely in terms of bringing about a more adequate mastery of these general tool skills as such, rather than in a specific mastery of certain definite materials.

remedying of defects in learning. In many ways the chief distinction between good instructional material and good remedial material is found in the time when it is used in the teaching process. If used for initial instruction it is teaching material. If used for the correction of difficulties arising from an incomplete mastery of the initial teaching it is remedial material. Whatever the function of the material, its nature depends upon the definite identification of the skills upon which effective achievement rests.

Arithmetic is one of the more definite tool subjects, and in general this subject matter has been systematically organized for teaching purposes. For years it has been recognized that success in addition depended upon a mastery of the basic addition facts. The same may be said of each of the four fundamental processes with whole numbers. It is somewhat significant, however, that it is only relatively recently that teachers have recognized that success in such work as long division is dependent upon a great many more skills than are involved in the mastery of the basic division facts. Long division calls for the accurate and effective use of skills in addition, multiplication, and subtraction, not to mention the skills which are usually recognized as being definitely division. Multiplication itself may involve the basic multiplication facts, the addition and multiplication involved in carrying in multiplication, and addition itself. Teachers are now coming to see that instead of being obliged to master only approximately four hundred simple facts in the four processes with whole numbers the pupil is called upon to master over four times as many. As Osborn[2] points out many teachers of arithmetic do not give the pupil very much chance to master the subject. Only a small portion of the specific facts needed for successful work are taught. It is assumed that the child himself should be able to figure out for himself that because 9+8 equals 17, it would be obvious to him that 39+8 equals 47. This the

[2] Osburn, W. J., "Corrective Arithmetic." Houghton Mifflin Company, 1925.

ordinary child does not do, and can not be expected to do without assistance. Furthermore it is often assumed that merely because the child knows individually each item or skill in each of the processes, he will be able to bring them together into their proper order and relationships in the working of a complex example. This again has been shown to be unsound, and we now know that we must aid the child definitely and directly in the process of putting together these separate skills in the solution of an example. This process is called synthesis in the later discussion in this chapter.

Recently a large amount of interest in the analysis of the field of arithmetic into its basic teachable and measurable skills has been created by the work of Thorndike,[3] Osburn,[4] Knight,[5] and others working in this field. One of the most complete catalogues of arithmetical skills selected for teaching, testing, and remedial purposes is the one on which the *Compass Diagnostic Tests in Arithmetic*[6] are based. A summarization of this analysis is presented here as an illustration of the extent to which such an analysis may be carried as well as to furnish a broad basis upon which to build remedial material in this field.

Arithmetic Skills

I. Fundamental Processes with Whole Numbers
   1. Basic Addition Facts
   2. Basic Subtraction Facts
   3. Basic Multiplication Facts
   4. Basic Short Division Facts
   5. Basic Vocabulary and Definitions of Arithmetic

[3] Thorndike, E. L., "The New Method in Arithmetic," Rand, McNally and Company, 1921. "The Psychology of Arithmetic," The Macmillan Company, 1922.

[4] Osburn, W. J., "Corrective Arithmetic." Houghton Mifflin Company.

[5] Knight, F. B., Ruch, G. M., and Studebaker, J. W., "Standard Service Arithmetics." Scott, Foresman and Company, 1927.

[6] Ruch, G. M., Knight, F. B., Greene, H. A., Studebaker, J. W., "The Compass Diagnostic Tests in Arithmetic." Scott, Foresman and Company, 1925.

6. Basic Rules of Arithmetic
7. Higher Decade Addition
8. Column Addition
9. Carrying in Column Addition
10. Harder Subtraction
11. Checking Errors in Subtraction
12. Borrowing or Carrying in Subtraction
13. Addition Used in Harder Multiplication
14. Carrying in Addition Used in Harder Multiplication
15. Complete Process of Multiplication
16. Short Division Involving Carrying
17. Multiplication, Addition, and Subtraction Used in Long Division
18. Complete Process of Long Division

II. Fundamental Processes with Fractions and Whole Numbers
1. Changing Fractions to Equivalent Forms
2. Finding Common Denominators
3. Reducing Fractions
4. Addition of Fractions and Mixed Numbers
5. Expressing Mixed Numbers as Improper Fractions
6. Fundamentals of Subtraction of Fractions
7. Reduction of Mixed Numbers
8. Cancellation in the Multiplication of Fractions
9. Reduction of Fractions and Mixed Numbers to Best Form in Answer
10. Multiplication of Fractions
11. Cancellation in Division of Fractions
12. Changing from Multiplication to Division Form
13. Fundamentals of Division of Fractions

III. Fundamental Processes with Decimals
1. Notation of Decimals
2. Changing Fractions and Mixed Numbers to Decimal Form
3. Changing Decimals to Fractions and Mixed Numbers
4. Fundamentals of Addition of Decimals
5. Fundamentals of Subtraction of Decimals
6. Place Values in Decimals
7. Pointing off in Multiplication of Decimals
8. Dividing Decimals by pointing off
9. Location of Decimal Points in Division
10. Changing Remainders to Decimal Form
11. Fundamentals of Division of Decimals

IV. Fundamental Processes with Denominate Numbers
1. Knowledge of Tables of Measure
2. Reducing in Denominate Numbers
3. Borrowing in Denominate Numbers
4. Addition of Denominate Numbers
5. Subtraction of Denominate Numbers
6. Multiplication of Denominate Numbers
7. Division of Denominate Numbers

V. Mensuration
1. Vocabulary of Mensuration
2. Mensuration of Plain Surfaces
3. Mensuration of Solids
4. Finding Areas and Volumes
5. Formulas Used in Mensuration

VI. Percentage
1. Fractional and Per Cent Relations
2. Decimal and Per Cent Relations
3. Expressing Areas in Per Cents
4. Fundamentals of Work in Percentage

VII. Interest
1. Vocabulary of Interest
2. Business Forms
3. Budgets
4. Computation of Interest
5. Computation of Discount
6. Use of Interest Tables

VIII. Problem Solving
1. Mastery of Fundamental Processes
2. Comprehension of Material Read in Problem
3. Knowledge of What Is Given in the Problem
4. Knowledge of What Is Called for in the Problem
5. Probable Answer to the Problem
6. Knowledge of Proper Processes to Use in Solving
7. Knowledge of Proper Order of these Processes
8. Recognition of the Correct Solution

The 70 or more distinct skills enumerated in this outline are not the only ones which may be identified, nor are they the only ones covered by the *Compass Diagnostic Tests* to which previous reference has been made. As a matter of fact something over 90 distinct items contributing to ability in the various aspects of arithmetic are covered in these

tests. Naturally the opportunities for the development of remedial exercises closely approximating these functions are correspondingly magnified as the analysis becomes more thorough.

A somewhat different but equally interesting and valuable list of specific unit skills in arithmetic is given in connection with the *Buswell-John Diagnostic Charts for Individual Difficulties in Arithmetic* to which reference was made in Chapter VIII. While the work habits listed in these charts represent the undesirable or error side of the skill in question they at the same time very definitely call attention to the corrective instruction needed. The following list of 28 work habits in addition represents the specific types of difficulties for which the examiner is instructed to watch in using this test. The list is reproduced with the permission of the publishers.

List of observed habits of work presented to guide the examiner using the Buswell-John Diagnostic Charts in Arithmetic

a1 Errors in combinations
a2 Counting
a3 Added carried number last
a4 Forgot to add carried number
a5 Repeated work after partly done
a6 Added carried number irregularly
a7 Wrote number to be carried
a8 Irregular procedure in column
a9 Carried wrong number
a10 Grouped two or more numbers
a11 Splits numbers into parts
a12 Used wrong fundamental operation
a13 Lost place in column
a14 Depended on visualization
a15 Disregarded column position
a16 Omitted one or more digits
a17 Errors in reading numbers
a18 Dropped back one or more tens
a19 Derived unknown combination from a familiar one
a20 Disregarded one column
a21 Error in writing answer
a22 Skipped one or more decades
a23 Carrying when nothing to carry
a24 Used scratch paper
a25 Added in pairs giving last sum as answer
a26 Added same digit in two columns
a27 Wrote carried number in answer
a28 Added same number twice

The main feature of the type of diagnostic and remedial work which may be built around the skills enumerated in this section has been repeatedly emphasized in the discussion in this and the preceding chapter. The reader should keep in mind at all times that pupils do not fail in a vague, general sense, or need remedial work of a vague and general type. Pupils' errors and failures are specific. The more exactly they can be located the more promptly they can be removed. Diagnostic tests based upon a satisfactory analysis of the skills which are essential to pupil mastery are for the purpose of locating these specific breakdowns. Related remedial units are to supply the properly organized drill needed to correct each specific breakdown.

Remedial exercises incorporating most of the desirable characteristics of such material can be developed by the classroom teacher, or can be secured in commercial form in certain school subjects. The preparation and use of such material to supplement available instructional devices should serve to increase greatly the efficiency of teaching particularly in the so-called tool subjects. It should be remembered that the most effective use of remedial material will follow the careful diagnosis of individual pupil difficulties by means of standard tests prepared for the purpose, and that the use of the tests without the accompanying remedial program is equally futile. Descriptions of a few of the available remedial exercises in arithmetic and some specific suggestions for the development of suitable drill exercises where commercial material is not available are presented in the following pages of this chapter.

II. Remedial Instruction in Arithmetic

127. Materials for Remedial Work in Arithmetic. An examination of available practice and drill material in the field of arithmetic reveals two somewhat distinctive types of material and uses of such material. General practice exercises designed to simplify the first learning and to aid in

maintaining a general mastery of the skill are by far the most common. Exercises designed to serve the more specific purpose of eradicating observed weaknesses and inadequacies in skills following initial instruction represent a more recent development in this field. The reader's attention has already been called to the main features of these two types of drill materials in Section 117 of Chapter VIII.

Materials for General Drill. A number of sets of practice exercises not specifically designed for remedial purposes are given in the following list in alphabetical order because of the varying instructional merit they possess:

Arithmetic Work Books (Ruch, Knight, Studebaker)
Scott, Foresman and Company, Chicago.
Clapp's Arithmetical Drills
Silver Burdett and Company, New York.
Courtis Standard Practice Exercises in Arithmetic
World Book Company, Yonkers-on-Hudson, N. Y.
Economy Practice Exercises in Arithmetic
Scott, Foresman and Company, Chicago.
Fowlkes-Goff Practice Tests in Arithmetic
The Macmillan Company, New York.
Hillegas' Drill Work in Arithmetic
J. B. Lippincott Company, Philadelphia, Pa.
Lennes' Work, Drill, and Test Sheets
Laidlaw Brothers, Chicago.
Thorndike Practice Books in Arithmetic
Rand McNally and Company, Chicago.
Work Book in Arithmetic (Myers)
Harter School Supply Company, Cleveland, Ohio.

Materials of Specific Remedial Nature. The arithmetic drill devices which are designed particularly for remedial purposes are not numerous. Illustrations and brief descriptions of representative material in this field are reproduced here in order to give the reader a more definite idea of the nature of such drills.

Osburn's Corrective Exercises in the Fundamentals of Arithmetic [7]

These corrective exercises consist of a series of graded drills in the four fundamental processes with whole numbers for use in grades two, three, and four. The usefulness of these exercises is confined to whole numbers only which, of course, somewhat limits their scope of use to the beginning elementary school grades. The general form of the material is indicated by the accompanying reproduction of Exercise 1 for the second grade. This material is reproduced with the permission of the publishers.

Specimen Exercise from Osburn's Corrective Exercises in the Fundamentals of Arithmetic

ADDITION—SECOND GRADE [8]

EXERCISE 1

5	3	1	3	4	6	7	0	2	5	0
2	5	4	6	0	1	2	4	2	1	5
1	2	2	0	7	1	0	2	4	7	8
7	4	6	3	0	5	6	1	5	1	0
1	1	2	2	0	3	6	1	0	3	6
0	2	5	7	9	0	0	6	2	2	2
4	5	4	0	3	4	1	3	0	4	1
4	3	1	7	1	3	1	3	1	2	3

[7] Osburn, W. J., "Corrective Exercises in the Fundamentals of Arithmetic." Houghton Mifflin Company, 1924.

[8] Copyright, 1925, by Worth J. Osburn; published by Houghton Mifflin Company, Boston. Reproduced with permission.

—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—
5	5	0	8	1	9	2	0	6	3	2
0	4	8	1	8	0	3	0	3	4	0
—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—

Time, 5 minutes.

Economy Remedial Exercises in Whole Numbers

In the field of commercial materials for remedial work in arithmetic the *Economy Remedial Exercises* probably lead in the detail and accuracy of analysis as well as in the extent of the remedial program available for use wholly or in part. In their ultimate form they will provide remedial units which will parallel closely the major skills in arithmetic enumerated on pages 235 to 237. At the present time there are 41 remedial cards for *Whole Numbers,* 41 cards for *Mixed Numbers and Fractions,* and 54 cards in *Problem Solving* all adapted to specially designed cards which permit the repeated use of the exercises.

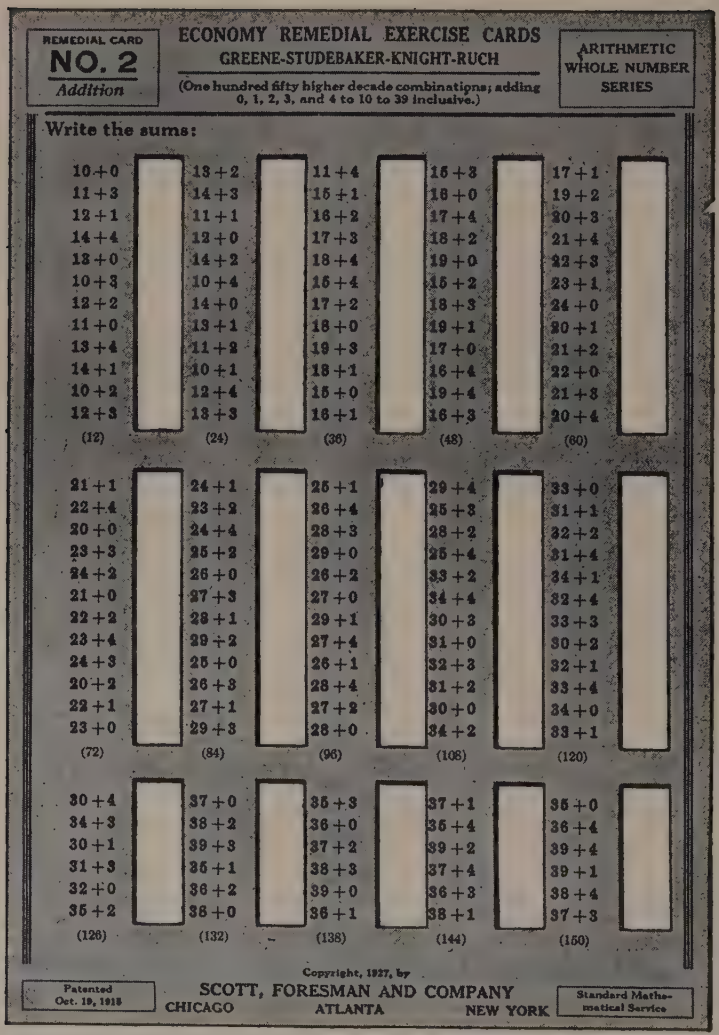
Some notion of the organization of this material within a given field of arithmetic may be gained from the accompanying table (Table 46) in which the proposed units of remedial drill designed to correct difficulties in the multiplication of whole numbers are shown in their relation to the basic skills to be developed.

The remedial exercise card reproduced on page 244 represents one of three cards furnishing drill on 450 of the higher decade addition facts. This particular card contains the 150 higher decade additions formed by adding zero to four, inclusive, to ten to thirty-nine, inclusive. The 450 higher decade facts represented by these three cards are practically all of the higher decade facts which work in ordinary column addition calls into play. Naturally many other cards in addition alone are necessary in order to synthesize properly the addition skills.

Table 46

SCOPE OF REMEDIAL DRILL UNITS IN WHOLE NUMBERS

For weaknesses in these Basic Skills	Use these types of Remedial Exercise Units
Basic Addition Facts	100 Addition Facts
Basic Subtraction Facts	100 Subtraction facts; easy combinations, no carrying or borrowing
Basic Multiplication Facts	100 Multiplication Facts
Basic Division Facts	90 Division Facts
Higher Decade Addition	450 Higher Decade Addition Facts
Column Addition	450 Higher Decade Addition Facts
Carrying in Column Addition	Introducing easy carrying
Harder Subtraction	Harder subtraction combinations introducing borrowing or carrying
Addition Used in Harder Multiplication	Higher decade addition facts
Carrying in addition used in Harder Multiplication	The 360 multiplication and addition combinations used in carrying in multiplication
Complete Process of Multiplication	Units introducing one-, two-, three-, and four-figure multipliers, and zero difficulties
Short Division with Carrying	The 360 short division combinations involving carrying
Addition, Subtraction, and Multiplication involved in Long Division	Previous units in these fields
Estimating Quotients	Units introducing estimation of apparent and non-apparent quotients
Complete Process of Long Division	Units introducing the complete process of long division one step at a time



REMEDIAL CARD
NO. 2
Addition

ECONOMY REMEDIAL EXERCISE CARDS
GREENE-STUDEBAKER-KNIGHT-RUCH
(One hundred fifty higher decade combinations; adding 0, 1, 2, 3, and 4 to 10 to 39 inclusive.)

ARITHMETIC
WHOLE NUMBER
SERIES

Write the sums:

10+0		13+2		11+4		15+3		17+1	
11+3		14+3		15+1		16+0		19+2	
12+1		11+1		16+2		17+4		20+3	
14+4		12+0		17+3		18+2		21+4	
13+0		14+2		18+4		19+0		22+3	
10+3		10+4		15+4		15+2		23+1	
12+2		14+0		17+2		18+3		24+0	
11+0		13+1		18+0		19+1		20+1	
13+4		11+2		19+3		17+0		21+2	
14+1		10+1		18+1		16+4		22+0	
10+2		12+4		15+0		19+4		21+3	
12+3		13+3		16+1		16+3		20+4	
(12)		(24)		(36)		(48)		(60)	
21+1		24+1		25+1		29+4		33+0	
22+4		23+2		26+4		25+3		31+1	
20+0		24+4		28+3		28+2		32+2	
23+3		25+2		29+0		25+4		31+4	
24+2		26+0		26+2		33+2		34+1	
21+0		27+3		27+0		34+4		32+4	
22+2		28+1		29+1		30+3		33+3	
23+4		29+2		27+4		31+0		30+2	
24+3		25+0		26+1		32+3		32+1	
20+2		26+3		28+4		31+2		33+4	
22+1		27+1		27+2		30+0		34+0	
23+0		29+3		28+0		34+2		33+1	
(72)		(84)		(96)		(108)		(120)	
30+4		37+0		35+3		37+1		35+0	
34+3		38+2		36+0		35+4		36+4	
30+1		39+3		37+2		39+2		39+4	
31+3		35+1		38+3		37+4		39+1	
32+0		36+2		39+0		36+3		38+4	
35+2		38+0		36+1		38+1		37+3	
(126)		(132)		(138)		(144)		(150)	

Copyright, 1927, by

Patented Oct. 19, 1918

SCOTT, FORESMAN AND COMPANY
CHICAGO ATLANTA NEW YORK

Standard Mathematical Service

Specimen Copy of Remedial Card in Arithmetic. Reproduced by permission of the publishers.

Economy Problem Solving Exercises

The development of skill in the solution of verbal problems is one of the more important goals of instruction in arithmetic. It is at the same time one of the more difficult skills to develop because of its complexity. The fact that only one set of drill exercises specifically designed to produce mastery in this field has been developed to date prompts the presentation of two further illustrations from this field. The sample reproduced on this page is a Process Analysis exercise from the "Economy Problem Solving Exercises."[9] The sample given on page 246 illustrates a set of exercises from this same series designed to train the pupil in determining *what is called for* in the problem, a skill which is clearly essential to the successful solution of the problem.

Specimen Exercises

No. 1

ECONOMY PROBLEM SOLVING EXERCISE CARDS[10]

GREENE-STUDEBAKER-KNIGHT-RUCH

Process Analysis—One-Step Problems

Directions: Read each problem carefully, but do not work it. Decide how you *would* work it. Would you add? Or Subtract? Or multiply? Or Divide? If you would add, write A in the hole beside the problem. If you would subtract, write S. If you would multiply, write M. If you would divide, write D.

The sample shows you how to write your answers. Study the sample carefully.

Sample: [S] *We have ten boys in our room at school. Four were absent today. How many were present?*

[ ] 1. My father bought 10 tons of coal at $11.00 a ton. How much did all of this coal cost?	7. How many "Big Chief" pencils that cost 5 cents each can I buy for 15 cents? [ ]
[ ] 2. How many "Junior League" baseballs at 75 cents each can I buy for $3.75?	8. From a class of 24 pupils 7 were absent. How many were present? [ ]

(*Twelve Exercises*)

[9] Greene, H. A., Studebaker, J. W., Knight, F. B., and Ruch, G. M., "Economy Problem Solving Exercises." Scott, Foresman and Company.

[10] From "Economy Problem Solving Exercises." Copyright, 1928 by Scott, Foresman and Company, publishers, Chicago.

No. 30

ECONOMY PROBLEM SOLVING EXERCISE CARDS

GREENE-STUDEBAKER-KNIGHT-RUCH

What Is Called for—One-Step Problems

DIRECTIONS: Read each problem carefully. Then read each statement (*a–e*) below the problem. Choose the *one* statement which tells *what is called for* by the problem; then write in the hole below the problem, the letter which stands for this statement. Ask yourself: "What would I find out by working this problem?"

1. *Mrs. Marvin bought silk for a dress at* **$1.90** *per yard. How much did Mrs. Marvin pay for four yards of the silk?*

(*a*) The kind of cloth Mrs. Marvin bought.
(*b*) The price Mrs. Marvin paid for the silk per yard.
(*c*) The total cost of the four yards of silk.
(*d*) The total cost of making the dress.
(*e*) The number of yards of silk required.

4. *What will it cost our class to buy baseball suits for twelve boys if each suit costs* **$9.50**?

(*a*) The number of baseball suits needed.....(*a*)
(*b*) The cost of one baseball suit.....(*b*)
(*c*) The number of boys on a baseball team.....(*c*)
(*d*) The total cost of the twelve baseball suits.....(*d*)
(*e*) The amount left after buying the suits.....(*e*)

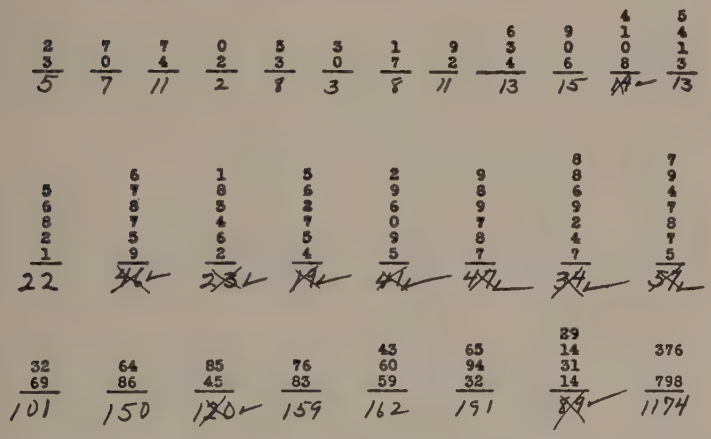
(*Six Exercises*)

128. Using the Remedial Materials in the Classroom. The use of remedial exercises in the classroom assumes that some type of reliable diagnosis of pupil weaknesses in arithmetic has been made. This diagnosis may be the result of making a careful analysis of the types of errors made by the pupil in his written work; it may be the result of careful observation on the part of the teacher; or it may be the result of the use of one or more of the types of individual or group diagnostic tests which have been discussed and illustrated in this book. However, it is not the *method* of the diagnosis but the *reliability* of the results that counts, for only through accurate diagnosis can the exact remedy be applied.

The location of individual pupil difficulty in arithmetic and the application of the correct type of remedial drill is illustrated in the case of the experience of Miss W., a fourth grade teacher, with Jimmie C., a student in her class. Jimmie C. has just entered Miss W's class from another school sys-

tem. His record shows that he has been ill a great deal, and as a result has been irregular in his school attendance. It is his arithmetic work which seems to have suffered most, however. In order to secure a more exact notion of the difficulties which he finds in his arithmetic work Miss W. pre-

JIMMIE C'S ARITHMETIC PAPER (PREPARED BY MISS W.)



pares for him a little test consisting of three types of examples.[11] Jimmie's test paper after Miss W. has corrected

[11] The steps taken by Miss W. in the preparation of this test paper for Jimmie C. are so interesting that a part of her work is reviewed here for the benefit of the reader interested in developing a similar type of test. First Miss W. prepared a sheet of cross-ruled paper and wrote the numbers from 0 to 47 along the left margin and the numbers 0 to 9 along the columns at the top of the paper as shown in the accompanying table. This sheet afforded Miss W. a basis for making a sampling of the basic and higher decade addition facts which would be representative. This table (Table 47) shows, for example, that the combination $0 + 3$ appears once in the test; the combination $7 + 8$ appears once, etc. It is by means of such a chart that the distribution of practice in texts, tests, and practice exercises is checked and controlled. The student will do well to check the distribution of practice provided in the test prepared by Miss W. It will be noted that this brief and simple test samples 59 of the 100 basic addition facts, and furthermore gives experience with 34 higher decade addition facts.

Table 47

CHART SHOWING DISTRIBUTION OF PRACTICE IN ADDITION TEST

(PREPARED BY MISS W.)

Number Added to	Number Added										
	0	1	2	3	4	5	6	7	8	9	
0				1				1			
1		1	1	1	1	1	1	1	1	1	
2	1			1	1		1			1	
3		1	1			1	1		1		
4		1		1	1	1		1		1	
5		1			1	1		1	1	1	
6	1		1		1	1	1			1	
7		1		1	1		1		1		
8	1	1		1	1	1	1	1			
9	1		1	1	1	1	1	1		1	(59)
10								1			
11			1				1				
12				1	1				1		
13							1		1		
14	1						1	1			
15								1	1		
16			1								
17						1					
18							1				
19											
20								1		1	
21									1		
22							1			1	
23		1									
24						1					
25											
26											
27					1						
28									1		
29			1					1			
30											
31									1	1	
32											
33											
34											
35											
36							1		1		
37											
38											
39										1	
40								1			
41											
42											
43											
44											
45											
46											
47						1					(34)

it is shown in page 247. Miss W. totals the number of each of the three types of examples he has worked correctly. From this she discovers that Jimmie knows the majority of the basic addition facts (11 of the 12 examples), but does not know the higher decade addition facts involved in doing column addition (only 1 of the 8 examples). Carrying in simple addition of more than one column, interestingly enough, does not bother him. (He worked 6 of the 8 examples right. The 2 he missed were undoubtedly due to carelessness since he worked harder examples in this set.) Miss W. now has a clue to one of the difficulties which Jimmie is having with his arithmetic examples. Her problem is now much simplified. She must drill Jimmie on the higher decade addition facts which are used in column addition. This she does and after a few weeks Jimmie is happy in his work in arithmetic which had previously been a nightmare to him. If space permitted many such cases might be mentioned as illustrative of the effective classroom use of remedial methods.

III. Summary

This chapter on the program of remedial teaching following diagnosis in arithmetic carries the discussion begun in the preceding chapter to a more concrete application. The meaning and purposes of remedial instruction in general are developed and illustrated in a specific field. As a basis for the evaluation of existing remedial materials in arithmetic and as a foundation for the development of new materials the specific skills in arithmetic are catalogued. Later these are used to show how remedial units must be developed paralleling these skills if corrective methods are to prove effective. Illustrations of a number of representative types of remedial units are included. One concrete illustration of the application of diagnostic and remedial methods by the classroom teacher in the solution of her instructional problem is given.

The reader will, of course, realize that even the most thorough understanding of the contents of this brief chapter

will not make him a master of the diagnostic and remedial techniques in arithmetic. Rather as a result he may have the beginnings of an appreciation of the problems involved and a better notion of how to proceed with the practical work. Only through such practical work, checked and changed continually in the light of honest self-criticism, will the reader become a master of diagnostic and remedial work in arithmetic.

A very comprehensive list of standardized tests and drill materials in arithmetic is given on pages 337–339 and 353–354 in the *Appendix* of this volume.

The problems involved in the application of diagnostic and remedial methods in the field of language are discussed in Chapter X.

SELECTED REFERENCES

Courtis, S. A., "Standard Practice Exercises in Arithmetic. Teachers Manual." World Book Company, Yonkers-on-Hudson, New York, 1916.

Greene, H. A.; Studebaker, J. W.; Knight, F. B.; Ruch, G. M., "Economy Remedial Exercises in Problem Solving in Arithmetic," Scott, Foresman and Company, Chicago, 1926.

Greene, H. A.; Studebaker, J. W.; Knight, F. B.; Ruch, G. M., "Economy Remedial Exercises in Arithmetic, Manual for Whole Number Series." Scott, Foresman and Company, Chicago, 1926.

Lennes, N. J., "Teachers Handbook for Work, Drill, and Test Sheets." Laidlaw Brothers, Chicago, 1924.

Morton, R. L., "Teaching Arithmetic in the Primary Grades." Silver Burdett and Company, New York, 1927.

Morton, R. L., "Teaching Arithmetic in the Intermediate Grades." Silver Burdett and Company, New York, 1927.

Osburn, W. J., "Corrective Exercises in the Fundamentals of Arithmetic." Houghton Mifflin Company, Boston, 1925.

Thorndike, E. L., "The New Methods in Arithmetic," Rand, McNally and Company, 1921.

CHAPTER X

DIAGNOSTIC AND REMEDIAL TECHNIQUES; LANGUAGE

The purpose of this chapter is to summarize the following points on the more important factors involved in the identification and correction of pupil difficulties in the field of language.

a. The social importance of language.
b. The place of formal language instruction.
c. The analysis of specific language skills, oral and written.
d. The scope of the diagnostic possibilities in language.
e. The possible scope of remedial work in language.
f. Some suggested types of remedial exercises in language.

129. Introduction. The importance of language in every day life naturally gives it a place of prominence among the instructional problems faced by the classroom teacher. It is surely significant that language which was one of the first subjects to be measured is one of the slowest to respond to analysis, diagnosis, and remedial treatment. Possibly this is due in part to the formal methods of instruction in this subject followed by teachers in spite of the rather significant amount of evidence against it.[1] As early as 1906 Hoyt[2] showed that when language ability was defined as *the ability*

[1] It is interesting to note that many people are willing to believe that training which improves an analytical skill like grammar will similarly improve a creative skill like composition. At the same time, these same people will not contend that the ability to write prose will carry with it corresponding ability to write poetry. Yet prose and poetry, both being creative, may be much more closely allied than grammar and either prose or poetry.

[2] Hoyt, Franklin S., "Studies in English Grammar." *Teachers College Record*, November, 1906.

to write a composition or to interpret literature, pupils in the seventh and eighth grades with no training in formal grammar did as effective work as those with two years of drill on formal grammar. Later Briggs [3] submitted evidence showing the failure of formal grammar to transfer to such skills as the statement of a definition, the application of a definition, or the ability to correct language errors. Other data on this question tend to lead to the conclusion that grammar as taught in these experiments is not a particularly effective instrument for giving a mastery of the most important socially useful language skills.

Students of the problems of the teaching of English are among the first to recognize the fact that in the minds of many, formal grammar is a discredited subject. Klapper,[4] for example, points out five distinct charges which may be brought against the current methods of teaching grammar.

1. "The old boast, 'Grammar teaches how to write and speak a language correctly,' has been disproven, not only by practical results observed in actual teaching but by a deeper analysis of the relation that exists between speech and grammar. . . . Exercises in grammar are essentially analytic; exercises in composition are creative and essentially synthetic; therefore ability in one of these forms of language study is not necessarily carried over to the other."
2. "In most classrooms, there is little or no relation between the courses in grammar and in composition for a given term."
3. "Grammar as outlined in many courses of study and in textbooks written for elementary schools abounds in sterile verbal subtleties."
4. "Grammar is still a memory subject rather than a rational study."
5. "No attempt is made to reach any degree of uniformity in the terminology; each book insists on its own system; and each author is a law unto himself."

These indictments of the teaching of formal grammar undoubtedly have some basis for existence. However, Klapper very correctly points out that they are weaknesses in present

[3] Briggs, T. H., "Formal English Grammar as a Discipline." *Teachers College Record*, September, 1913.

[4] Klapper, Paul, "Teaching English in Elementary and Junior High Schools." D. Appleton and Company, New York, 1925.

practices rather than weaknesses which are necessarily inherent in the subject itself. In defense of grammar, when taught from a properly organized course of study and by pedagogically sound methods of teaching, Klapper submits the following values:[5]

1. It is a guide to correct speech, but does not guarantee the development of it.
2. It is a means of developing power in the use and analysis of verbal relationships.
3. It is a valuable asset in literary interpretation and in comprehensive reading.[6]
4. It is justified on the ground of social expediency. "Many facts are taught, not because they have intrinsic worth but because they form part of that knowledge stock that society expects its citizen of culture to possess."
5. It is a necessary preparation for future studies, with due allowance being made for the fact that only a relatively small percentage of pupils pursue studies which depend upon it.

130. A Language Consciousness. It is not the purpose of this discussion to attempt to settle the matter of the relative emphasis which formal grammar teaching should have in the development of language skill. The important thing is that the skills be developed. In this discussion language skill is considered to mean facility in the use of the proper language habits and forms essential to effective intercommunication. Such a point of view makes reasonably clear the problems of the language teacher. Language skills arise as do other specific skills, through the proper exercise of the desired habits. It goes without saying that proper exercise

[5] These statements of the values of grammar are not quotations from Klapper except as noted, but are adaptation of his more extensive discussion of the values in his "Teaching English in Elementary and Junior High Schools," referred to on page 252 of this chapter.

[6] On this point Klapper refers to the study of Hoyt and concludes that Hoyt's findings are exactly what one would expect under the circumstances. "They prove that proficiency in any of these three branches of the study of English (composition, grammar, and literary interpretation) is no index of the proficiency that will be attained in the other two." He further points out that only in awkward language situations will a knowledge of grammar be expected to aid in literary interpretation.

is possible only when proper identification of the habits has taken place. One can not indulge in exercises until he knows what to exercise. Hence the habits upon which language skill depends must be identified, and much carefully constructed instructional and drill material must be provided. This in itself will serve two useful purposes. First, the use of good language drill material insures that the pupil will have experience in making the correct response to selected language situations either with or without the assistance of such formal grammar instruction as may be applied. In the second place, the use of such material sets up in the pupil's mind an attitude toward language error. A definite consciousness toward language errors must be developed. When a person becomes sensitive to language errors, and an expression such as "he don't" causes a reaction similar to that created in a cat confronted by a strange dog, he is on the way to rapid improvement in his language habits, without regard to the way in which this "consciousness" may be created.

I. Identification of Specific Language Abilities

131. Analysis of Language Skills. It must be evident from what has been presented in earlier chapters in this book that an accurate analysis of the underlying skills in language is necessary before any significant program of diagnostic and corrective work can be undertaken. In the past, certain general language abilities have been identified for measurement purposes; such as language usage, grammar, composition, etc. The various scales for the measurement of composition ability; notably, the *Hillegas Composition Scale,*[7] the *Thorndike Extension of the Hillegas Scales,*[7] the *Nassau (Trabue) Supplement to the Hillegas Scale,*[7] the *Trabue Language Scales,*[7] and many others, are representative of these general language measures. In more recent years, however, there has been an effort to reduce language in general to its more elementary or basic skills. These attempts seem to be in the right direction for practice on narrow skills which

[7] Published by Bureau of Publications, Teachers College, Columbia University, New York. See *Appendix* for more complete list.

together make the total ability in language have been found helpful. Some of these devices for the measurement of the more specific functions of language ability are *Briggs' English Form Test,*[8] *Franseen Diagnostic Tests in Language, Kirby Grammar Tests, Pressey Diagnostic Test in English Composition, Wilson Language Error Tests,* and many others.

These more specific measures of language ability are undoubtedly the outgrowth of such constructive thought on the problems of English teaching as that given by the *National Joint Committee on Reorganization of English in Secondary Schools* under the chairmanship of Dr. Hosic.[9] The following summary of the aims and abilities supposed to result from English instruction (Sections 130 and 131) is an adaptation and restatement of these outcomes as enunciated in the report of this committee, supplemented by certain observations of the authors of this book.

The task of clarifying the statements and purposes of language (English) instruction and of analyzing and identifying the basic language skills is not one which will be successfully accomplished by any one person. A great many different analyses must be made by many different people under different conditions and from different points of view before all of the basic skills can be catalogued. At the outset it must be recognized that there are many conditions under which language functions. There is undoubtedly a language of impression, or comprehension, as well as a language of expression. The former is the aspect of language which is given particular attention in reading instruction. The latter, the language of expression, is the phase usually meant by the term "language ability," and is the phase which receives special attention in language and English instruction.

132. Oral Language Skills. Language as a means of verbal expression appears in two main forms, oral and written. Success in the use of oral language depends in the first place

[8] See the *Appendix* of this book for information relative to the publishers of these and other tests in the field of language.

[9] Hosic, J. F., "Reorganization of English in Secondary Schools." Bureau of Education Bulletin No. 2, 1917.

upon the individual's ability to choose, arrange, and enunciate his words in such a manner as to affect his hearers as he intends. In order to guarantee success in the operation of these skills the student must be given training and practice in connected thinking and talking under audience conditions. In this training emphasis must be placed on the development of a pleasant speaking voice, a gracious attitude, a clear enunciation of words, an avoidance of common language errors, care in the selection of words, a careful selection and organization of ideas, and skill in the clothing of his thoughts in the proper words that he may affect his hearers as he intends by leading their thinking along prescribed channels.

133. Written Language Skills. The problem of written language takes a threefold form. The first involves the formal or mechanical factors such as writing, spelling, punctuation, form, and general appearance. The second treats of certain grammatical factors such as common errors in language form, sentence structure and form, etc. The third is concerned with the more subtle elements of composition, the rhetorical factors involving the questions of choice of words, the quality of interest innate in the material, and the logical organization of the subject matter both within the sentence and the larger units. In the first two phases of the problem of written language the factors are more generally uniform in their manner of affecting readers. However, there is greater difficulty in predicting the effect of the third phase on the reader. These mechanical and grammatical elements constitute in a way the raw material of written expression. The rhetorical factors are the results of the manner in which these raw materials are put together. They are the factors which make for appeal, originality, and distinctiveness in written expression. The mechanical and grammatical factors are relatively tangible, objective, and measurable. The rhetorical factors are more elusive, more difficult to identify and measure, but offer nevertheless significant opportunities for the development of remedial materials in the field of language.

II. Diagnosis of Language Disabilities

134. Diagnosis of Oral Language Disabilities. Considerable progress in the identification of oral language disabilities and in the development of remedial exercises in oral expression has already been made by investigators in the field of speech. Prominent among these investigators are Smiley Blanton, formerly of the Minneapolis Child Guidance Clinic; May K. Scripture, School of Education, New York University; Lee Edward Travis, Director of University of Iowa Speech Clinic; Richard C. Borden; and Alvin C. Busse; each of whom has made concrete contributions to this field. Blanton[10] recognizes four fundamental speech disorders and proposes specific remedial training for their correction. These are: (1) delayed speech; (2) oral inactivities; (3) letter substitution; and (4) stuttering, which includes stammering. In connection with each of these types of disorders he points out suggestions and methods of correction which are sufficiently pertinent to the problem of oral language expression to warrant some reproduction here.

Delayed Speech. Four periods of speech development in the child are noted: (1) the period of reflex and instinctive cries; (2) the period of language sounds without association of ideas; (3) the period of language sounds with association of ideas, but without sentence formation; and (4) the period of sentence building. Blanton holds that a child should be able to get what he wants by means of oral language by the time he is fifteen months of age, and that if he cannot talk with some degree of facility by the time he is two and a half years of age he should be given careful mental and neurological examinations. The four causes he lists for delayed speech are: (1) lack of mental development; (2) lack of necessity for speech (i.e., why go to the trouble of learn-

[10] Blanton, Smiley, *Problems and Methods in the Correction of Defective Speech.* In "Speech Training and Public Speaking for Secondary Schools," a report of a special committee of the National Association of Teachers of Speech, compiled and edited by A. M. Drummond, Chairman of the committee. The Century Company, New York, 1925.

ing to talk if one can get everything he wants by simpler methods); (3) continued illness with extreme malnutrition during infancy; and (4) infantile neuroses.

Oral Inactivities. This classification covers a large group of speech defects resulting from inactivity of the lips, jaw, tongue, or soft palate. In general two causes are recognized: (1) some organic weakness caused by disease in infancy; and (2) some injury to the nervous system causing indirectly a weakness or paralysis of the speech organs.

Letter Substitution. Letter substitution or lisping covers a large class of sound substitutions, found generally under two types. In one, the subject always substitutes one sound for another. In the other he is able to produce all of the sounds but places them in the wrong order. From the point of view of remedial work these two types are very different. If the child is unable to make the proper sound he may require corrective work in phonetics to help him to master it, while if he merely misplaces the sounds, it is mainly a matter of ear-training in the recognition of the sounds.

Stuttering and Stammering. These defects, which may be described as "a break in the rhythm of speech arising from an inability to make the muscular co-ordinations necessary for speech," are both basically psychological. To quote Blanton further: "The real cause of stuttering is the fear, partly conscious and partly unconscious, of meeting the group. The child fears to meet the group, but he also has a desire to do so. These tendencies to flee and to meet the situation come into conflict, and we have a compromise, in which we have neither good speech nor absence of speech, but broken, inhibited, stuttering speech." Exercises in phonetics, breathing, etc., are mostly useless for the difficulty is not in the speech mechanism. Training the child in general bodily relaxation in order to aid in reducing the emotional strain under which the stuttering child seems to labor is one of the most helpful remedial practices in this field.

It is obviously impossible in this brief volume to consider in further detail the problems of remedial work in the correction of oral language disabilities. Enough has been given

however, to make the student or classroom teacher aware of the importance of this aspect of language ability. The responsibility of the school for aiding the child handicapped in this way should now be somewhat more clear. The teacher especially interested in the further study of this interesting phase of language will find the *Quarterly Journal of Speech Education* a mine rich in material of a theoretical as well as a practical nature.

135. Diagnosis of Difficulties in Written Language. In the discussion of language skills presented in Sections 131 and 132, basic skills contributing to verbal expression are grouped under the general headings of the oral and written forms. This is a convenient and satisfactory classification for the purpose of contrasting the two major types of verbal expression, but it seems inadequate when considered from the point of view of the complete identification and analysis of the specific underlying skills upon which verbal expression depends. For this more exacting purpose a classification based upon the units of language form; namely, the word, the sentence, the paragraph, the composition unit, and certain general mechanical factors, is superior. In order to present a more concrete notion of the types of abilities called into play at each of these levels of language skill the accompanying detailed outline [11] is given.

Diagnostic Outline of Language Skills

A. Words—Skill in the spelling, choice, use, and definition
  1. Spelling—ability to spell certain socially useful words
  2. Choice of Words
     a. synonyms
     b. antonyms
     c. exact word for the place
     d. variety
     e. meaningful words
     f. minimum number of words

[11] This outline is reproduced with permission from an unpublished study by Mr. H. L. Ballenger.

3. Correct Usage
   a. verbs
   b. pronouns
   c. modifiers
   d. nouns
4. Use of Dictionary

B. Sentences—Skill in the use, form, structure, and organization

1. Form
   a. complete, coherent, unified sentences
   b. initial letter capitalized
   c. correct terminal punctuation
   d. variation in beginning
   e. variation in length
2. Structure
   a. simple (declarative, interrogative, exclamatory)
   b. subject and predicate
   c. variety in structure
   d. language usage
      (1) slang and foreign expressions
      (2) faulty expressions
      (3) double negative
3. Organization
   a. logical sequence of ideas

C. Paragraphs—Skill in the form, structure, and organization

1. Form
   a. indentation
   b. initial and terminal line length
   c. length
2. Structure
   a. unity
   b. coherence
3. Organization
   a. outline
   b. logical sequence of ideas

D. Letter Writing—Skill in

1. Business
2. Social
3. Informal notes
4. Formal notes

E. General Mechanical Factors—Skill in control of

1. Capitalization
2. Punctuation

3. Margins
4. Handwriting
5. Abbreviations
6. Hyphenations
7. Spelling

In spite of the detail of this outline and the number of specific language skills enumerated the reader will immediately recognize certain significant weaknesses. Many of the skills are identified only in a very general way. The recognition of choice of words as a significant language skill is approximately equal to stating that addition is an important skill in arithmetic. Much more definite information is necessary before all of the details of a constructive program of language improvement can be developed. Just as it is necessary to identify the 100 basic addition facts, the 450 or more higher decade addition facts, etc., in arithmetic it is necessary for us to identify the words which have the greatest social usefulness in language situations. Much excellent work has been done on the problem of determining a minimal spelling (writing) vocabulary based on social utility.[12] Similar work must be done from the standpoint of language situations. Until this is accomplished workers interested in the development of diagnostic exercises in written language must turn to other sources for valid test and drill materials.

136. Diagnostic Exercises in Written Language. An examination of the outline of language skills given on pages 259–260 reveals a large number of opportunities for the analytical measurement of these abilities. One of the most recent and extensive attempts at the development of a diagnostic language test which has come to the attention of the authors is that prepared by the author of the accompanying outline of language skills. A few exercises from parts of this test are reproduced here as illustrations of the types of measurement attempted by the tests.

[12] Horn, Ernest, "A Basic Writing Vocabulary." University of Iowa Monograph in Education.

A Diagnostic Language Test[13]

TEST 1. SPELLING

(Fifty Words)

Directions to Pupil: Some of the words below are spelled correctly and some are not. *You are to write each word correctly on the line at the right of the word.* Begin with the first word and go down the first column; then do the same with the second column. Be sure to spell each word as it should be spelled, whether or not it is spelled correctly in the list.

1. acept	accept	**6.** perchase	_____
2. recive	_____	**7.** carefuly	_____
3. arange	_____	**8.** standerd	_____
4. wemon	_____	**9.** shown	_____
5. gentelmen	_____	**10.** Saterday	_____

[13] These exercises are reproduced with permission from an experimental form of a diagnostic language test prepared by Mr. H. L. Ballenger.

TEST 2. WORD MEANINGS

PART A. SYNONYMS

(Thirty Exercises)

Directions to Pupil: In each exercise below, one of the four numbered words is *very similar in meaning* to the first word in the exercise. *Find this word and place the number that is in front of it on the line at the right.* The sample is answered correctly. Answer the other exercises in a similar way.

Sample: Queer.... (1) love.... (2) odd.... (3) run.... (4) man.... *2*

1. Small	(1) true	(2) real	(3) light	(4) little	_____
2. Sure	(1) near	(2) certain	(3) finish	(4) free	_____
3. Said	(1) please	(2) rain	(3) stated	(4) mind	_____
4. High	(1) view	(2) tall	(3) free	(4) price	_____
5. Wise	(1) mind	(2) sell	(3) sensible	(4) plan	_____

PART B. OPPOSITES

(Thirty Exercises)

Directions to Pupil: In each exercise below, one of the four numbered words is *opposite in meaning* to the first word in the exercise. *Find this word and place the number that is in front of it on the line at the right.* The sample is answered correctly. Answer the other exercises in a similar way.

Sample: Far....	**(1) field....**	**(2) paper....**	**(3) near....**	**(4) house....**	*3*
1. Begin.....	(1) sure.....	(2) finish.....	(3) tired.....	(4) help.....	
2. Stop.....	(1) continue....	(2) enter.....	(3) choose.....	(4) delight.....	
3. Strange.....	(1) familiar.....	(2) please.....	(3) permit.....	(4) quit.....	
4. Often.....	(1) stayed.....	(2) began.....	(3) seldom.....	(4) worried.....	
5. Open.....	(1) room.....	(2) long.....	(3) close.....	(4) place.....	

TEST 3. CHOICE OF WORDS

(Fifty Exercises)

Directions to Pupil: In each of the following sentences, decide which of the numbered words is correct to use in the sentence, then *write the number that is in front of the correct word on the line at the right of the exercise.* **Do not guess.** The sample is answered correctly.

Sample: He was a (1) **queer** (2) **funny** man..... *1*

---

1. The weather is (1) **awful** (2) **very** cold..... _____
2. He came to school (1) **too** (2) **to** early..... _____
3. (1) **Can** (2) **May** I go with you?..... _____
4. You cannot (1) **teach** (2) **learn** a lazy boy..... _____
5. I spent (1) **two** (2) **too** dollars for that hat..... _____

TEST 4. CORRECT USAGE

(Fifty Exercises)

Directions to Pupil: In the following sentences, certain words are printed in heavy or black type. *On the line at the right of each exercise write "yes" if the indicated word or group of words is used correctly. If the word or group of words is used wrong, write "no" on the line.* The sample is answered correctly. **Do not guess.**

Sample: The boy has **ran** a mile..... *No*

(The word *ran* is used wrong, so the word "*No*" has been written on the line.)

1. **Was** there any ripe apples on your tree?.....
2. Mary **did** her work well.....
3. I thought you **was** Tom.....
4. Where **were** you going when I saw you yesterday?.....
5. Tom **done** what he was told to do.....

TEST 5. GRAMMATICAL FORM RECOGNITION

(To be omitted in Grades 4 and 5)

(Fifty Exercises)

Directions to Pupil: The word which has been printed in heavy or black face type in each of the sentences below, may be a **noun, pronoun, adjective, adverb, verb, conjunction, preposition,** or **interjection.** *Write which of these each word is, at the right of each sentence in the blank space provided for that purpose.* You may use the following abbreviations: noun (*N*), pronoun (*Pr*), adjective (*A*), adverb (*Adv*), verb (*V*), conjunction (*Con*), preposition (Prep), interjection (*I*). The sample is answered correctly.

Sample: The **man** over there is my uncle..... *N*

1. The **children** in the school were very happy.....
2. The house in which they lived was **large** and white.....
3. The children were glad when school **began**.....
4. How **beautiful** the flowers were!.....
5. A new house has been **built** near us.....

TEST 6. FAULTY EXPRESSIONS

(Fifty Exercises)

Directions to Pupils: Some of the sentences below are not good ones because they contain slang phrases, faulty expressions or double negatives. *Opposite each exercise that you think is good write G. Opposite each sentence that you think is bad, write B.* The sample below is answered correctly. **Do not guess.**

Sample: I didn't do nothing ..... *B*

1. Didn't you see nobody on the corner? .....
2. I am not sure whether he is here .....
3. They ain't going to Chicago today .....
4. I am sure that you sort of want to go .....
5. He has said nothing all morning .....

TEST 7. SENTENCE SENSE

(Twenty-five Exercises)

Directions to Pupil: Some of the following statements are complete sentences and some of them are not complete. *On the line at the right of each exercise that* **is** *a complete sentence, write the letter S, and opposite each exercise that* **is not** *a complete sentence write the letter N.* **Do not guess.** The sample below is answered correctly.

Sample: On the way to the store ..... *N*

1. We helped Rose with her lesson .....
2. One day as Roy was feeding the chickens .....
3. At the school .....
4. The small boy cried .....
5. Paul, the best talker of all .....

TEST 8. SENTENCE STRUCTURE

(Twenty Exercises)

Directions to Pupil: Below are some sentences with numbers placed under various parts of each sentence. *At the right of each exercise, write* **first** *the number that is under the* **subject** *of that sentence and then at the right of this number, write the number that is under the* **predicate,** as in the sample below:

Sample: My brother has a football *1 2*

1 2 3

---

1. The Eskimo baby wears a little fur suit . . . . . _____

1 2 3

2. The hunter was sitting alone in his hut by the fire . . . . . _____

1 2 3 4 5

3. How fast the train goes! . . . . . _____

1 2 3 4

4. Where are the pencils that were on my desk? . . . . . _____

1 2 3 4 5 6

5. There has been a wolf among the sheep . . . . . _____

1 2 3 4

TEST 9. SENTENCE ORGANIZATION

(Twenty Exercises)

Directions to Pupil: This test is given to see how well you are able to arrange groups of words into sentences. The sample below is answered correctly. *Answer the other exercises in a similar way.*

Sample: (1) the paper, (2) the man, (3) reads *2 3 1*

---

1. (1) on the ground, (2) acorns, (3) lie ..... _____
2. (1) a top, (2) had, (3) the little boy ..... _____
3. (1) ran, (2) the fox, (3) into the woods ..... _____
4. (1) sing, (2) the children, (3) they, (4) heard ..... _____
5. (1) this winter, (2) to skate, (3) learn ..... _____

TEST 10. CAPITALIZATION AND PUNCTUATION

PART A. CAPITALIZATION

(Fifty Exercises)

Directions to Pupil: The sentences below are correct, except that certain capital letters are missing from some of the words. *On the line at the right of each exercise, write the word that should begin with a capital letter.* The sample is answered correctly.

Sample: We moved into our new house in march..... *March*

1. I saw jack in the store.....
2. When will you and i go to school?.....
3. We went to the country for christmas.....
4. I read the poem, "obedience".....
5. Last saturday was a very pleasant day.....

PART B. PUNCTUATION

(Fifty Exercises)

Directions to Pupils: In each of the following exercises, some form of punctuation is missing. *On the line at the right of each statement, write the word which comes just before the missing punctuation mark and then show what has been omitted.* The sample below is answered correctly.

Sample: The boy saw an apple on the tree ..... *tree.*

---

1. Will you go with me .....

2. Where is Louis R Smith now? .....

3. Dr Jones called on us last night. .....

4. Hurrah for the flag of America .....

5. Yes we will wait for you. .....

TEST 11. PARAGRAPH STRUCTURE

(Twenty Exercises)

Directions to Pupil: In each of the following exercises, there is one sentence that does not belong in the paragraph. Each sentence has been numbered. *On the line at the end of each exercise write the number of the sentence which does not belong with the other sentences in the paragraph.*

---

1. (1) Once upon a time Father Fox, Mother Fox, and three little baby foxes lived in a hole in a tree. (2) They were pretty red foxes and as sly as sly could be. (3) Old Mr. Wolf knew where to find the fattest sheep and just how to catch them..... _____

2. (1) There was once a shepherd boy who tended sheep at the foot of a mountain. (2) So the shepherd boy lost all his flock. (3) It was a lonely place for him, and he often wished for company and a little fun..... _____

3. (1) The prince returned home and was very happy. (2) One night there was a terrible storm. (3) The wind blew and blew; it thundered and lightened, and the rain poured down in torrents.... _____

4. (1) The boys worked and worked. (2) The girls helped the Pilgrim mothers. (3) They helped to sew and make the clothes for the men and boys..... _____

5. (1) The Tin Soldier held himself as stiff as he could. (2) No one could say that he moved an eyelid. (3) The boat dashed over the edge of the drain into the deep canal..... _____

TEST 12. PARAGRAPH ORGANIZATION

(Twenty Exercises)

Directions to Pupil: The following exercises are given to test your ability to arrange the sentences of a paragraph in their proper order. In these exercises the sentences are not all in proper order. *At the right of each exercise place the numbers of the sentences in the order in which they should appear.* The sample is answered correctly.

Sample: (1) While we were there we saw the men plow the corn and make hay. (2) We left home Saturday morning on our trip to the country and returned in the evening..... *2 1*

---

1. (1) The wind blew and blew; it thundered and lightened, and the rain poured down in torrents. (2) One night there was a terrible storm.....

2. (1) Early next morning Mr. Robinson, Fritz and I started off to see the island. (2) Fan was left at home to take care of Mrs. Robinson and the children.....

3. (1) At last it filled up to the brim with water; it must sink. (2) The boat swirled round, and round, and round.....

4. (1) It did not look like a brick house, for it had been covered on the outside with boards. (2) In a little town not far away from Boston stood an old brick house.....

5. (1) When it storms, the wind does not blow so hard in a river and the waves are not so high as on the ocean. (2) A ship can usually find a safe harbor by sailing a short distance up into a river...

III. Remedial Instruction in Language

137. Language Diagnosis as Basis for Remedial Instruction. A comparison of the skills which the language test exercises presented in Section 136 are designed to measure, with the content of the diagnostic outline presented on pages 259–261, reveals the fact that these twelve types of language ability are among the most important measurable skills thus far identified. The specific type of diagnosis which such a detailed diagnostic test should furnish presents some very definite possibilities for the development of remedial instruction in language.

In the following sections (Sections 138 to 141 inclusive) illustrations of specific types of remedial exercises in language are presented as suggestions to the teacher interested in the further development of such drill material. Certain of these exercises have been subjected to sufficient classroom use to prove tentatively their remedial value. Others are presented because of their close parallel with the types of exercises which seem to be able to reveal the presence or absence of the indicated skills in a pupil. Theoretically they should be valid remedial drills. Unfortunately in the majority of these exercises their value has not been determined experimentally. Much experimental work remains to be done in this field.

138. Remedial Exercises in Language; Spelling. Spelling is one of the important factors limiting the successful use of words in written expression, and as such is properly classified among the general mechanical factors in language. (See outline on page 259.) In general, materials for the correction of mechanical difficulties in written expression are relatively simple to develop. Work of this type has been undertaken by almost every classroom teacher, although the majority of it is done as specific instruction in the subject and is not correlated with language work as such.

In addition to the best known methods which may be applied to the teaching of spelling there are distinct possibilities

in the use of so-called proof-reading exercises in language for the development of a critical or "editorial" attitude toward misspellings in written material. This is a concrete case of the type of "consciousness" described in Section 130 of this chapter. The following illustrations may be suggestive of material which can be used in this way.

Spelling Exercise

Certain words in this paragraph are misspelled. Read it carefully and underline each misspelled word.

All men are buy natsur equaly free and independent, and hav certen inherent wrights, the enjoiment of lyfe, libertie, with the means of accuiring and posessing preparty, and pursueing and obtaneing hapiness and saftie.

Spelling Exercise

In the following list of twenty words some are misspelled. Write each misspelled word correctly at the right of the word.

annual	_____	alowed	_____
beleive	_____	cource	_____
resurve	_____	volume	_____
minite	_____	terible	_____
copies	_____	deside	_____
judje	_____	offiser	_____
notise	_____	chief	_____
faver	_____	acker	_____
phone	_____	receve	_____
explane	_____	colect	_____

139. Remedial Exercises in Punctuation. Dictation exercises based upon content material or narrative material suited for the grade provide suitable drill on punctuation and the use of capitals. The following illustration is taken from *Briggs' Form Test Alpha*:[14]

[14] Reproduced with permission. Copyright 1920, by Teachers College. Published by Bureau of Publications, Teachers College, Columbia University, New York City.

Punctuation and Sentence Form Exercise

Read each group of words so as to get the meaning. Then in the proper places put capital letters, apostrophes, necessary commas, periods, and question marks.

1. birds sing
2. Where is the fire
3. In april the apple trees were in bloom many motorists stopped to admire them.
4. The boys hat was torn but his clothes were neat.
5. the girls are sewing this morning

. . . . . .

19. In all the history of the village there are a few stories of genuine pathos one of these concerns a young kentucky boy left an orphan by the sudden and accidental death of both parents they were on a visit to their parents at the time of the tragedy of which we have heard.
20. Here and there on the dusty shelves of Higgins & Browns dilapidated general merchandise store the boys found cans of salmon and corned beef but their appetites were not whetted by thoughts of a meal of tinned food without a morsel of anything else to go with it.

140. Remedial Exercises in Form and Appearance. The social importance of form and appearance particularly in letter writing, one of the most frequent social uses of written composition, places a premium on such skills in this field. Exercises of the following types are suggested as useful drills on letter form.

Letter Form Exercise

Arrange, capitalize, and punctuate properly the following items which go to make up the heading and salutation of a letter:

gentlemen
236 erie avenue
columbus city mo
july 18 1931
the sanford and morris company

(Use your own home city and address in the heading of the letter.)

Letter Form Exercise

Write as I dictate the following items in the proper form for a business letter:

Your home address is 140 Grand Avenue Boulevard.
Your home town is Bluffton, Pennsylvania.
The date is (*give present date*).
You are writing to Williams and Burke, Attorneys-at-law, whose address is Sprague Building, 10th and Ferry Sts., Long Beach, California.

Use the proper salutation (gentlemen) for a letter of this sort.

141. Remedial Exercises in Correct Language Usage. Many excellent remedial exercises for the correction of common language errors may be devised once the most frequently appearing errors are known. Material such as is utilized in Test 8 of the *Stanford Achievement Tests,*[15] or in the *Charters Language Tests,*[16] *Kirby Grammar Tests,*[17] *Pressey-Conkling Grammar Tests,*[16] and *Wilson's Language Error Test*[15] meets most of the common language error situations necessary in a test. Naturally these test exercises are not presented in such form as to give any opportunity for a pupil to secure practice on the correct response for the purpose of fixing it in mind as a desirable language habit. The development of many supplementary exercises paralleling sections of the tests noted above should prove very helpful to the classroom teacher. The following samples will serve to illustrate suitable types of exercises.

Exercise in Finding Language Errors

A number of mistakes have been made in writing this story. See if you can find the mistakes and correct them.

John come to town yesterday. It is show day. He eat dinner with me and then I takes him to the show. After we had

[15] World Book Company.
[16] Public School Publishing Company.
[17] Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.

saw the show John have started to walk home but a man takes him home in his car. If I had knowed we could have rode I would have went along.

Exercise in Finding Language Errors

In each exercise there are two forms in parentheses, one is correct, and the other is incorrect. Mark out the word that is incorrect.

1. (Whom) (Who) did you send?
2. He told Mary and (me) (I) a story.
3. I could (of) (have) gone along.
4. The book (don't) (doesn't) belong to me.
5. Either your work or mine (is) (are) wrong.

142. Remedial Exercises in Sentence Structure. Exercises of the following sort will afford effective remedial drill for children having difficulties in sentence structure.

Sentence Structure Exercise [18]

In each of the following groups of sentences there are some statements which are not well expressed. Place a cross (×) in front of each such statement to show that it is not a good sentence.

I asked her the name of the book she was reading.
He was glad to leave, for he was tired and sick of the place, for he had made no friends.
She told me the names of her sister and kitty.
Mary had a good position. Which she left.
She is happier than I.

143. Remedial Drill in Choice of Words. A great deal of difficulty is encountered by youthful (and adult) writers in the matter of choice of words. Such errors as "He does his work good" are not uncommon. In many cases it is a matter

[18] This exercise and the one following are adapted from material incorporated in the Course of Study in Language used in the University of Iowa, Elementary School, Iowa City.

of choosing a word with a more exact shade of meaning, and in many cases it is a matter of knowing the form of a given word to use. This is particularly true in the field of adjectives and attributive nouns. Exercises such as the following giving drill in choosing the correct compared form of certain adjectives will be found helpful in cases where specific diagnosis reveals this type of weakness.

Exercise in Choice of Words

Read the following exercises and fill in the missing word needed to complete the meaning of the sentence. In each the missing word in an exercise is a form of the underlined word in the sentence.

1. New York is a large city. It is ———— than Chicago; in fact it is the ———— city in the United States.
2. I want to buy a good fountain pen. I want a ———— one than that. Show me the ———— one you have.
3. John is a mischievous boy. He is ———— ———— than Bob. He is the ———— ———— boy in school.
4. This is a bad storm, much ———— than the one we had last week, but not the ———— one this year.

144. Drill on Language Organization. Relatively little is known concerning the possibilities of training children in language organization. Investigation has shown that children who do well on the *Greene English Organization Test* tend to write compositions of higher quality than those who make relatively lower scores. Some preliminary experimentation on a limited number of cases also shows that children respond to training on exercises similar to these test exercises and that this training seems to be a factor in the accompanying improvement of written productions. This experimentation, however, has been confined to organization within the sentence. In an objective sense practically nothing is known of these aspects of language organization. What evidence there is tends to show that exercises such as the following may have a certain remedial value.

Exercise in Organization in Sentences [19]

The groups of words printed below do not make sense. How can they be arranged so as to make a good sentence? Read the groups of words in the exercises below and directly below the exercises write the words arranged in a good sentence.

a top, had, the little boy

Answer: The little boy had a top.

1. sing, the children, they, heard

2. wanted, to go home, him, the dog

3. and made them, into the palace, their own, took them, they

4. of rejoicing, to have, all wanted, they were, a great feast, that they, and so thankful, so happy

Exercise in Paragraph Organization

The sentences in the following paragraphs are not arranged in the best order. On the line below each paragraph, write the numbers of each of these sentences in such order that the arrangement will be improved.

(1) One man found, that until he put toads in his greenhouse, he could not keep insects from eating some of his flowers.
(2) Sometimes men keep toads in their greenhouses.

Answer: (*2, 1*)

Exercise in Paragraph Organization

(1) These stories are not true. (2) Isn't the toad a wonderful animal? (3) Although the toad is a greedy eater, he can live a long time without food or water. (4) However, toads have been known to live two whole years without either food or

[19] Adapted from material presented in Greene, H. A., "Measurement of Linguistic Organization in Sentences," University of Iowa Study in Education, No. 4, Vol. II, July, 1923.

water. (5) Stories are told of toads that have been shut up in rocks for many years and have lived without food or water.[20]

Answer: (*3, 5, 1, 4, 2*)

In addition to these illustrations of more or less informal types of language exercise material an extensive list of commercial drill materials in the language field is given on page 354 in the *Appendix* of this volume. It will be noted that drill material of a definitely remedial nature is somewhat more limited in language than is the case in arithmetic and certain other elementary school subjects.

IV. Summary

The program of diagnosis and remedial teaching in language outlined in this chapter is based upon the conception of language as including all of the skills and processes by means of which the intercommunication of ideas, thoughts, and emotions is possible. Language ability as the means of verbal expression is divided into the fairly well recognized fields of oral and written expression. Many of the skills involved in these two fields are common, and to a certain extent the same corrective procedures may be utilized. However, written language represents the most fruitful field for the application of analytical, diagnostic, and remedial methods up to the present time.

The possible scope of the diagnosis of language abilities is dependent upon the clarity with which the language skills themselves are identified. A catalogue of certain recognized skills in the language field is given as a basis for the development of diagnostic test exercises, and as a basis for suggestions for the development of corrective drill exercises in language.

[20] Material adapted and used with permission from Horn-McBroom "Learn to Study Readers," Book 2, Ginn and Company, 1924.

An extensive list of standardized tests and drill materials in the field of language is given on pages 341–342 in the *Appendix* of this volume.

Chapter XI takes up the application of diagnostic and remedial methods in the field of reading.

SELECTED REFERENCES

Blanton, Smiley, *Problems and Methods in the Correction of Defective Speech* in "Speech Training and Public Speaking for Secondary Schools," a report of a special committee of the National Association of Teachers of Speech, compiled and edited by A. M. Drummond, Chairman. The Century Company, New York, 1925.

Clark, M. G., "Language in Use," *Teachers Manual.* Public School Publishing Company, Bloomington, Illinois, 1924.

Fletcher, John M., "The Problem of Stuttering." Longmans, Green & Co., New York, 1928.

Greene, H. A., "Measurement of Linguistic Organization in Sentences." University of Iowa Studies in Education, No. 4, Vol. II, July, 1923.

Hosic, J. F., *Reorganization of English in Secondary Schools; Report of Joint Committee on;* Bureau of Education Bulletin No. 2, 1917.

Klapper, Paul, "Teaching English in Elementary and Junior High Schools." D. Appleton and Company, New York, 1925.

Quarterly Journal of Speech Education. Official organ of the National Association of Teachers of Speech. Edited by John Dolman, Jr., University of Pennsylvania, Philadelphia.

Wohlfarth, Julia H., "Self-Help Methods of Teaching English." World Book Company, Yonkers-on-Hudson, New York, 1925.

CHAPTER XI

DIAGNOSTIC AND REMEDIAL TECHNIQUES; READING, AND OTHER SUBJECTS

The purpose of this chapter is to summarize the following important points involved in the identification and correction of instructional difficulties in reading and certain other school subjects:

a. The educational and social importance of reading.
b. The major objectives of reading instruction.
c. The problems of diagnosis in reading.
d. The scope of remedial possibilities in oral reading.
e. The scope of remedial work in reading of the work-study type.
f. Illustrations of types of remedial material in reading.
g. Diagnosis and remedial work in certain other school subjects.

145. Educational and Social Importance of Reading Ability. One of the most important recent changes in classroom instruction is the tendency to enrich the course of study by means of the enlarged reading opportunities offered to children. Instead of confining the pupils' reading to a single text book covering only a limited amount of information on any given subject the progressive school provides for and demands a wide range of reading activity on the part of the pupils. Furthermore, it is now being recognized that the solution of most classroom problems in the modern school requires the skillful use of books as sources of information. When viewed from this point of view reading is something more than merely the rapid comprehension of printed symbols and the

memory and organization of the materials read. It is also the ability to use books and libraries as efficient sources of information. This tendency to treat reading as a most important tool of learning has resulted in establishing a very close relation between reading [1] and practically every other school activity. As a means of gaining information and pleasure it is essential in every content subject such as history, geography, science, literature, and arithmetic.

In recent years there has also developed a keener appreciation of the importance of intelligent reading in society at large. The investigations of Gray [2] show that reading is an indispensable means of "familiarizing adults with current events, with significant social issues, with community and national problems, and with American institutions, ideals, and aspirations. It is also essential in attaining vocational efficiency, in broadening one's range of information, and in seeking pleasure and profit during leisure hours." It is true that "of the making of books there is no end." The mass of printed matter which the average adult must read and evaluate, even within the limits of his own fields of interest, is stupendous. This situation makes the development of a high degree of reading skill in our schools all the more essential.

The necessity for a high level of reading ability on the part of high school and college students is more readily realized when it is recognized that approximately ninety per cent of the vast bulk of facts they are supposed to master are obtained from books, or at least as a result of reading. This is, of course, an estimate, but the truth lies not so far from this figure. The real significance of the matter is seen in the fact that there is overwhelming evidence of a generally low level of reading ability on the part of these individuals who must use it so consistently. Book, in a recent study of this

[1] Gates, Arthur I., "The Improvement of Reading," The Macmillan Company, New York, 1927.

[2] Gray, W. S., "Importance of Intelligent Silent Reading." *Elementary School Journal*, January, 1924.

problem,[3] says: "It is clear why many college students fail to do satisfactory work in their college courses. Some are so deficient in their ability to get points made by the authors they read and from whom they must get the facts to be learned that we have little reason to expect success from them until this deficiency is successfully made up."

As a result of a program of special instruction given to 54 students under his direction Dr. Book reports in this same article that the reading efficiency of these individuals was improved 102 per cent during a semester. Their ability to master their assignments was increased from 60 per cent to 97.3 per cent efficiency. Some individuals improved their reading efficiency as much as 250 per cent. As a result of this and other supplementary investigations Dr. Book concludes "that the reading ability of college freshmen should be accurately determined; that special remedial instruction should be given in a special orientation or "how to study" course. Other recent data of a similar nature indicate that the reading abilities of adults can be markedly improved through systematic use of properly devised drill. In a recent study[4] Averill and Mueller report that sixteen adult subjects were able practically to double their rate of reading without decreasing their comprehension as a result of three forty-minute practice periods per week for approximately three months. It is interesting and significant that these results compare closely with the findings of O'Brien in his experiments with Illinois school children.[5]

The results of such investigations serve to reveal at least two important facts. First, the abilities of mature students and adults to apply their reading skills effectively are not in keeping with our assumptions. Second, the situation is not hope-

[3] Book, W. F., "How Well College Students Can Read." *School and Society*, Vol. XXVI, August 20th, 1927; pages 242–248.

[4] Averill, L. A., and Mueller, A. D., "The Effect of Practice on the Improvement of Silent Reading in Adults." *Journal of Educational Research*, Vol. 17; No. 2; February, 1928; pp. 125–129.

[5] O'Brien, J. A., "Reading, Its Psychology and Pedagogy." The Century Company, New York, 1926.

less, for the evidence shows that given proper diagnosis and a systematic application of properly constructed instructional and drill materials a significant amount of improvement in reading ability can be brought about in a relatively short time. The teacher's problem is therefore clear. The remaining sections of this chapter set forth in logical order steps which will be found helpful in the solution of the problem.

I. Identification of Major Reading Abilities

146. Major Objectives in Reading Instruction. The most common classification of reading objectives in the past has been to place them under the headings of oral and silent reading abilities. Recently it has seemed advisable further to cross-section these two types in reading skills and evaluate them in terms of their use in typical life situations. The life situations under which reading is usually done may be conveniently grouped into two types, depending on the attitude of the reader. These are:

1. Reading for pleasure, or reading of the leisure type.
2. Reading for information, or reading of the work-level type.

It must be clear that no hard and fast lines can be drawn between these two levels of reading activity since an individual may shift from one type to the other without realizing the change. However, the distinction between the type of situations under which reading ordinarily takes place affords a convenient basis for the analysis and location of specific reading skills around which to develop diagnostic and remedial materials.

The accompanying outline of the field of reading is adapted from an outline of reading skills prepared by Horn and McBroom,[6] and supplemented from a functional analysis

[6] Horn, Ernest, and McBroom, Maud, "A Survey of the Course of Study in Reading," University of Iowa Extension Bulletin No. 99, February, 1924, College of Education Series No. 3.

presented by Yoakam [7] on pages 54 to 56 of his "Reading and Study," an excellent treatment of the problem of securing more effective study methods through the development of better reading habits. A number of the uses reported by Gray as a result of his analysis [8] of reading activities of students in the University of Chicago High School, are also utilized in this outline.

A Functional Analysis of Reading

A. Reading of Leisure Types
- I. Oral Reading
  1. entertainment of others
  2. self-entertainment
  3. appreciation of beauty of expression
  4. practice on expression
- II. Silent Reading
  1. self-improvement
  2. self entertainment
  3. vicarious experience
  4. appreciation
  5. practice

B. Work-Types of Reading (major portion of reading time)
- I. Oral Reading
  1. entertainment of others
  2. instruction of others
  3. acquisition of new modes of expression
  4. extension of vocabulary
- II. Silent Reading (Major aspects of work-types of reading)
  1. Examples of specific situations in which one reads silently
     a. to secure the facts on which judgments may be based

[7] Yoakam, G. A., "Reading and Study," The Macmillan Company, New York, 1928.

[8] Gray, W. S., "The Relation between Study and Reading," *Proceedings of the National Education Association,* Vol. LVII, 1919; pages 580–86.

b. to secure information on the essential conditions of a problem to be solved
c. to verify a fact or an opinion
d. to secure a basis for action
e. to secure general information in desirable fields
f. to acquire information for specific use
g. to evaluate material as to its applicability to a specific case
h. to secure a basis for the formulation of an opinion from data or statements
i. to analyze the essential elements of an argument
j. to enlarge one's vocabulary
k. to acquire more effective modes of expression
l. to master technical and contextual meanings of words
m. to acquire a basis for a critical questioning attitude on a controversial issue
n. to provide a basis for the understanding of a specific situation
o. to satisfy the desire to discover new problems and answer them

III. Essential Skills Involved in Work-Study Types of Reading, and Suggested Methods of Acquiring Them

1. Skills and Abilities	2. Methods[9]
a. Skill in recognizing new words.	*a.* Methods suggested in manuals of standard method readers.
b. Ability to locate material quickly:	*b.* —
(1) Knowledge of and ability to use an index.	(1) Children learn alphabet, find words in alphabetical arrangement, find words in an index, find answers to questions by use of index, make an index for a book which has none.

[9] Quoted with permission from Horn-McBroom, *op. cit.*

(2) Ability to use a table of contents.	(2) Assign lessons by title; find lessons in table of contents; find authors in table of contents.
(3) Ability to use the dictionary.	(3) Handled in detail in Rice, O. S., *Lessons on the Use of Books and Libraries,* Rand, McNally and Company.
(4) Ability to use library card files.	(4) Handled in detail in Rice, O. S., *Lessons on the Use of Books and Libraries,* Rand, McNally and Company.
(5) Ability to use reference material.	(5) Handled in detail in Rice, O. S., *Lessons on the Use of Books and Libraries,* Rand, McNally and Company.
(6) Ability to use keys, tables, graphs, etc.	(6) Practice in interpreting maps, tables, etc., in geographies, etc.; make graphs or tables to illustrate problems.
(7) Ability to skim.	(7) Certain topic given, children skim to find material about it, to find answer to a question, a sentence which proves point, etc.
c. Ability to comprehend quickly what is read:	*c.* Use flash cards with words, phrases, sentences, requiring action responses; find answers to questions; answer true and false statements; follow directions; completion sen-
(1) Rhythmic and rapid eye movements.	
(2) Absence of lip reading.	

(3) Knowledge of meaning.	tences; labelling objects; matching words and objects, or words and pictures.
d. Ability to select and evaluate material needed.	*d.* Same as "c"; also practice noticing date of publications, name of author; choose from several statements one which answers question best.
e. Ability to organize what is read: (1) To summarize. (2) To assign topics to proper order or place. (3) To discover related material. (4) To outline.	*e.* Practice giving subjects to a paragraph; choose paragraph which answers questions; find everything which bears on a topic; put related words together; give a summary; outline a lesson or page.
f. Remembrance of material read:	*f.* Analysis of how to memorize quickly; practice in it; points in an article memorized; reports made without notes.
g. Knowledge of sources.	*g.* Practice using common sources: dictionary, encyclopedias, yearbooks, magazines, readers' guides, etc.; make lists of sources.
h. Attitude of attacking reading with vigor.	*h.* Hold every child responsible for jobs assigned; tie up reading with project work; show each child his needs through testing.
i. Attitude of proper care of books.	*i.* Lessons on how books are bound; what ruins bindings; how to keep books clean, how to make them endure; who pays for the books, etc.

The foregoing outline of the major objectives of reading instruction affords a useful basis for the evaluation of present instructional emphasis as well as a valuable list of criteria for the validation of diagnostic and remedial devices in reading.

147. Typical Defects in Reading. The solution of the problem of developing satisfactory and effective remedial materials in reading is dependent upon the accurate identification of specific causes of reading failure. As a result of a careful survey of the literature, Dr. W. S. Gray has listed the following fourteen outstanding causes,[10] any one of which may be the cause of difficulty in reading, and therefore may become the basis for a remedial program.

1. Inferior Learning Capacity
2. Congenital Word Blindness
3. Poor Auditory Memory
4. Defective Vision
5. Narrow Span of Recognition
6. Ineffective Eye Movements
7. Inadequate Training in Phonetics
8. Inadequate Attention to the Content
9. Inadequate Speaking Vocabulary
10. Small Meaning Vocabulary
11. Speech Defects
12. Lack of Interest
13. Guessing versus Accurate Recognition
14. Timidity

It must be apparent that while this catalogue of reading difficulties is somewhat less specific in its character than the skills enumerated on pages 289 to 292 of this chapter, it nevertheless opens up a field of genuine possibilities for the development of remedial and instructional materials. A number of illustrations of types of such materials are given in Sections 150 and 153 of this chapter.

[10] Gray, W. S., "Remedial Cases in Reading: Their Diagnosis and Treatment." Supplementary Educational Monographs, No. 22, June, 1922, University of Chicago Press.

148. Oral versus Silent Reading. An examination of the major aspects of remedial work in reading indicates that there are two angles from which it may be considered. In the first place remedial instruction may be begun in the oral reading field. Judd, Gray, and others defend this point of attack on the problem on the ground that it enables the teacher to start with the child on a level at which he already has some mastery, that is, the oral language level. Others believe that on account of the large proportion of reading time spent in the work-type of silent reading this field should receive the special emphasis. There is merit on both sides of the question, undoubtedly. It is true that the child does come to school with a fairly adequate oral vocabulary which in a great many ways affords the natural approach to reading. On the other hand, it is also true that such an approach tends to place too large an emphasis on the pronunciation of words and too little upon their meaning when encountered in silent reading situations. The transfer from the emphasis on oral reading (pronunciation of words) to silent reading (comprehension of meaning of words, sentences, and paragraphs) must be made at some point in the child's experience. Accordingly, a great many teachers hold that the place to start the emphasis on silent reading is at the beginning. Some foundation for this belief is seen in the results obtained by many teachers who place the emphasis on the development of silent reading skills at the outset.

II. Diagnosis and Remedy in Oral Reading

149. Diagnosis of Oral Reading Difficulties. The outline of skills in the field of reading presented on pages 289 to 292 afford an effective framework around which to build instruments for the diagnosis of specific reading difficulties. This section and the one which follows consider the problems of diagnosis as related to the oral reading fields and the work-study types of reading respectively.

The *Gray Standardized Oral Reading Paragraphs*[11] probably represent the most highly developed tests of oral reading ability. While they were not designed specifically for diagnostic purposes the fact that they are used as individual tests and not as group tests makes them rather useful diagnostic instruments for the teacher of the lower elementary school grades. They are standardized for use in the first to the eighth grades, but the shift in the emphasis from oral to silent reading in the upper grades decreases the significance of their use in the higher grades. Paragraphs 1, 6, and 12 of the test are reproduced here with the publishers' permission as an illustration of the character of the material.

Gray's Standardized Oral Reading Paragraphs

1[12]

A boy had a dog.
The dog ran into the woods.
The boy ran after the dog.
He wanted the dog to go home.
But the dog would not go home.
The little boy said,
"I cannot go home without my dog."
Then the boy began to cry.

6

The part of farming enjoyed most by a boy is the making of maple sugar. It is better than blackberrying and almost as good as fishing. One reason why a boy likes this work is that someone else does most of it. It is a sort of work in which he can appear to be very industrious and yet do but little.

[11] Gray, W. S., "Standardized Oral Reading Paragraphs," published by Public School Publishing Company, Bloomington, Illinois.

[12] There are 12 of these paragraphs arranged in increasing order of difficulty. The tests are administered individually by having the pupils read the paragraphs under time control until a certain number of errors per paragraph are made. Rate and quality of reading are then combined by means of a table of values provided for the purpose.

12

The hypotheses concerning physical phenomena formulated by the early philosophers proved to be inconsistent and in general not universally applicable. Before relatively accurate principles could be established, physicists, mathematicians, and statisticians had to combine forces and work arduously.

The plan of recording the number and kinds of errors made by the pupil in reading the *Gray Oral Reading Paragraphs* makes the results somewhat diagnostic from the oral reading standpoint. The accompanying quotation from the directions for giving the tests gives a more exact idea of the results. (See section 3b.)

DIRECTIONS FOR GIVING THE ORAL-READING TESTS

1. The pupils should be tested individually in a quiet place where they will be free from distraction and where the remainder of the pupils to be tested will not hear the reading.
2. When everything is in readiness to begin, hand the pupil a copy of the standardized paragraphs and give the following directions: "I should like you to read some of these paragraphs for me. Begin with the first paragraph when I say 'Begin.' Stop at the end of each paragraph until I say 'Next.' If you should find some hard words, read them as best you can without help and continue reading." Pupils above the fourth grade should begin with paragraph 4. If two or more errors are made in this paragraph, ask the pupil to read the preceding paragraphs also. In case pupils in the first two grades hesitate several seconds on a difficult word, pronounce it for the pupil and mark it as mispronounced.
3. While the pupil is reading, record two sets of facts in regard to the reading: the time required to read each paragraph and the errors made.

   (*a*) The time record is secured by noting the exact second at which the pupil begins reading a paragraph and the time when he completes it. The number of seconds required to read the paragraph should be recorded in the margin to the right of the paragraph.

(*b*) In order to illustrate clearly the character of errors and the method of recording them, the following paragraph is inserted:

The sun pierced into my large windows. It was the opening of October, and the sky was a dazzling blue. I looked out of my window down the street. The white houses of the long, straight street were almost painful to the eyes. The clear atmosphere allowed full play to the sun's brightness.

If a word is wholly mispronounced, underline it as in the case of "atmosphere." If a portion of a word is mispronounced, mark appropriately as indicated above: "pierced" pronounced in two syllables, sounding long *a* in "dazzling," omitting the *s* in "houses" or the *al* from "almost," or the *r* in "straight." Omitted words are marked as in the case of "of" and "and"; substitutions as in the case of "many" for "my"; insertions as in the case of "clear"; and repetitions as in the case of "to the sun's." Two or more words should be repeated to count as a repetition.

It is very difficult to record the exact nature of each error. Do this as nearly as you can. In all cases where you are unable to define clearly the specific character of the error, underline the word or portion of the word mispronounced. Be sure you put down a mark for each error. In case you are not sure that an error was made, give the pupil the benefit of the doubt. If the pupil has a slight foreign accent, distinguish carefully between this difficulty and real errors.

4. Each pupil should be allowed to continue reading until he makes seven errors in each of two paragraphs.

The analysis of the individual pupil's record gives a very concise picture of his oral reading difficulties. It will be noted that in these oral reading exercises no attention is paid to the degree of comprehension with which the material is read. The measurement of comprehension lies somewhat beyond the purpose of this test. Here the purpose is the determination of the efficiency with which words are recognized and pronounced in context.

In the *Gray Oral Reading Paragraphs* the purpose is the determination of the efficiency with which words are recog-

nized and pronounced in context. Gates proposes the comparative use of these tests with one of his own, the *Gates Pronunciation Tests,*[13] for the purpose of determining how far the pupil depends upon the contextual relationships of the words in the passage to make his attack on the word. He states that[14] "if a pupil does appreciably better on the Gray's Passages than on the Gates' Words, it is highly probable that the context has been fruitfully utilized. If the age or grade scores are approximately the same, it is probable that the context is utilized to an average degree. If the score is appreciably better on the Word Pronunciation Test, it is probable that the pupil makes relatively little use of the context—a deficiency that certainly should be remedied." A reproduction of the 100 words comprising the *Gates Pronunciation Test* is given on this and the following page.

Gates Graded Word Pronunciation Test

Name.....Grade....Age....Years....Months.....
School.....Date.....Total Score.....
Examiner's Name.....Remarks.....

FORM I

so	we	an	is	do	
as	go	us	at	or	10
the	not	did	can	hen	
how	may	out	son	net	20
king	here	door	ball	came	
grow	late	hear	east	year	30
every	about	after	broom	child	
paper	blind	peach	climb	point	40
window	family	bridge	lonely	scratch	
perhaps	plaster	harvest	servant	frighten	50

[13] Gates, Arthur I., "A Test of Ability in the Pronunciation of Words." *Teachers College Record,* November, 1924.

[14] Gates, Arthur I., "The Improvement of Reading." The Macmillan Company, New York, 1927.

passenger	wander	eighty	counter	shepherd	
interest	chocolate	motion	citizen	elegant	60
dispute	portion	continue	mansion	brilliant	
conductor	brightness	punishment	guardian	restaurant	70
intelligent	construct	impatient	protection	temperature	
position	profitable	instrument	reverence	astonishment	80
irregular	schoolmaster	manufacture	revolution	unnecessary	
lamentation	community	examination	intelligence	national	90
satisfactory	illustrious	countenance	congratulate	preparation	
superstition	affectionate	substantial	philosopher	treacherous	100

The purpose of this test is, of course, to determine how well the pupil can recognize and pronounce words when presented outside of their contextual setting. As is apparent from the general arrangement of the words it is a power test. The score is expressed in terms of the number of words pronounced correctly by the pupil. The test is standardized for both age and grade groups.

150. Remedial Drills for Oral Reading Difficulties. In the belief that the use of exercises designed to produce skill in both types of reading abilities is desirable a few illustrations of both kinds of drills are presented in this section and the one which follows (Section 153). In the field of oral reading a number of the drill exercises used in language training will be found effective (See Section III, pages 280 to 283 of Chapter X).

Dearborn, Gray, Buswell, Huey, and many others, studying the problem of how to improve oral reading, have noted that there is a marked relationship between the rate and quality of children's reading and the control they have over their eye-movements in reading. The meaning of eye-movements may be readily understood by anyone who will take a position closely in front of and directly in the range of vision of a person engaged in reading. The observer will note that the reader's eyes do not move regularly and systematically forward as the reading progresses but that the

movements are interspersed with pauses or fixation periods. It is during these pauses that the images of the words or groups of words are secured. Carefully conducted laboratory experiments reveal the fact that good readers make longer sweeps with the eyes, take in larger units of words, pause for much shorter periods, and rarely retrace material once covered. Superficially, these facts seem to indicate that training the pupil in the development of proper eye-movement habits is an effective remedial procedure. However, the question constantly arises whether the records of difficulty in eye-movements are indications of causes of reading difficulties or whether they are merely the results or expressions of difficulty in word recognition as such. If they are primary causes of reading difficulties then remedial treatment directed to the correction of improper eye-movements should be productive of improvement in reading. If they are indications of reading difficulties, only then is it likely that training on eye-movements as a direct means of improving reading ability would be futile. Gates in his "Psychology of Reading and Spelling,"[15] pages 56 ff., reports that he found practically no improvement in ability to read resulting from remedial work of this type. In a later discussion of this same problem Gates[16] concludes that improper eye-movements are probably the evidences of other types of reading disability which may be treated specifically. With the removal of the causes lying back of ineffective eye-movements the treatment of eye-movements as such becomes unnecessary. On the whole this seems to be the most hopeful way of looking at the problem since a great many teachers are qualified to administer types of remedial treatments which may be applied in the classroom but only a few have the technique or equipment for training the pupil in more effective eye-movements.

[15] Gates, Arthur I., "Psychology of Reading and Spelling." Teachers College Bureau of Publications, Columbia University, New York.

[16] Gates, Arthur I., "The Improvement of Reading." The Macmillan Company, 1927.

In his "Remedial Cases in Reading,"[17] Gray himself recognizes this practical aspect of the problem of remedial instruction in reading and suggests a number of excellent exercises designed to overcome specific difficulties in reading, many of which are indicated by the ineffectual eye-movements of the pupil. Three of these exercises are reproduced here as types of material adapted to remedying certain types of oral reading difficulties.

Exercises to Increase Accuracy of Recognition[18]

1. Words which a pupil failed to recognize accurately while reading were used in sentences at the end of each period in order that he might associate them with their meaning. The words which repeatedly caused difficulty were then typewritten on cards and used in quick-perception drills, by presenting them as rapidly as they were recognized. Such words as *again, want, been, does,* and *heard* were frequently emphasized. As soon as a pupil was able to recognize a word readily, drill on it was discontinued. New words were added to the list as difficulties were encountered.
2. Words which a pupil confused because of their similarity in form were emphasized in drill exercises. These words included such groups as *thought, though,* and *through, there* and *where, then* and *when, now* and *how,* and *has, had,* and *have.* The words were used in sentences before they were presented in quick-perception drills. If unusual difficulties were encountered, words which were similar in form were presented together so that their differences could be studied.
3. Pupils who recognized isolated words accurately frequently made errors in recognizing the same words in phrases and sentences. In order to overcome this difficulty a word, such as *there* was written on the board in several phrases or short sentences and the pupil was given opportunity to study them deliberately. As soon as he was able to recognize these

[17] Gray, W. S., "Remedial Cases in Reading; Their Diagnosis and Treatment." Supplementary Educational Monograph, No. 22, University of Chicago Press, 1922.

[18] Designed to remedy defects number 5, 6, and 13 in Gray's list given in Section 147, page 293 of this chapter. Reproduced with the permission of the author and publishers. *Op. cit.*

phrases readily they were typewritten on cards and presented in quick-perception drills.

Exercises to Increase Span of Recognition [19]

To aid a pupil in increasing his span of recognition, phrases were printed on cards and presented to him rapidly in flash-card, or quick-perception drills. At first, short phrases, such as *one day, every time, asked she, he said, very fast, over there,* and *as though,* were presented. These phrases were taken from the lessons which he was reading and the list included those which occurred most frequently. As soon as the pupil was able to recognize short phrases, longer and more difficult phrases were used, such as *again and again, over the hill, around the house, wherever he could, pretty little things, thought to himself, as fast as he could, while I am gone,* and *as much as he wanted.*

Exercises to Increase Ability in Word Recognition [20]

When a pupil had difficulty in recognizing and pronouncing words because he failed to associate sounds and their symbols, he was given some training in phonetics. If he was unable to pronounce a word, such as *fed,* he was given a familiar word, such as *bed,* and was asked to name others ending in *ed.* If he failed to recognize the familiar element *ed* other words, such as *red* and *led,* were supplied. If the initial consonant caused difficulty, other words beginning with *f* were pronounced for him and written on the board. If he had difficulty with the vowel sound, he was given lists of words containing short *e,* such as *met, set, get,* and *hen, when, den.* After several different endings and initial consonants had been used with short *e,* a chart containing such words as *let, pen, tell, fled,* and *bless* was presented to test the pupil's ability to recognize short *e* in various combinations. After all short vowel sounds or all long vowel sounds had been given, similar charts were prepared and used as a final test of recognition. This method prevented a child from depending entirely on so-called "families" in recognizing words. Initial consonant blends, vowel digraphs, and phonograms were taught in the same way.

[19] Gray, W. S., *Op. cit.*

[20] Designed to remedy defects numbers 7 and 13 in Gray's list in Section 147, page 293 of this chapter.

In the cases of many pupils certain rules of pronunciation were given in addition to the exercises which have just been described. The rules which proved most helpful follow.

1. When there is only one vowel in a word or syllable, that vowel is usually short as in *can, met, hit, not,* and *cup.*
2. When a one-syllable word ends in *e,* the *e* is silent and the preceding vowel is long, as in *lame, hope,* and *write.*
3. When there are two vowels together in a word or syllable, the first one is usually long and the second one is silent, as in *beat, rain,* and *coat.*

Lack of space forbids the inclusion of additional exercises of this type. Those interested in attacking the problem of oral reading deficiencies will find ample material presented in Gray's monograph, "Remedial Cases in Reading, Their Diagnosis and Treatment," and in Gates' "The Improvement of Reading." By practice in the use of diagnostic reading devices and by training themselves in careful observation of their pupils teachers can become adept in the detection of particular reading difficulties. Furthermore, they will soon find that with practice they can become proficient in the art of building drill exercises. The sample exercises selected from a great many suggested by Gray, Gates, and others will be found valuable guides in the preparation of such materials.

III. Diagnosis and Remedy of Silent Reading Difficulties

151. Diagnosis of Work-Study Type of Reading; Elementary School Level. The emphasis given to the work study types of reading in the list of skills given on pages 289 to 292 indicates something of the importance of the type of reading in relation to the total reading field. The reason why some pupils fail (in arithmetic, for example) is not entirely due to ignorance of the basic facts, or to lack of mental ability to understand the explanations, but rather to sheer inability to read. In fact, one of the best ways to improve work in many other school subjects is to make a

drive on the work-type of reading ability. A recognition of this has caused makers of tests in reading to turn their attention in this direction in recent years. A number of excellent reading tests which provide diagnostic information within certain limits are available. Among these tests are the *Detroit Reading Tests, Gates' Silent Reading Tests, Haggerty Silent Reading Examinations, Sigma* 1 *and Sigma* 3, *Sangren-Woody Reading Tests,* the *Stanford Achievement Test,* etc.[21] The *Gates' Primary Reading Tests* for grades one and two, the *Gates' Silent Reading Tests* (Types A, B, C, and D) for grades three to eight, and the *Sangren-Woody-Reading Tests* for grades four to eight are the most recent and provide the most discerning analysis of the basic reading skills.

Typical exercises from three of these tests are reproduced here as illustrations of the type of diagnosis possible through the use of carefully constructed and valid reading tests.

[21] The publishers and distributors of these tests will be found on page 348 of the *Appendix.* A much more complete list of reading tests is also given in the *Appendix.*

GATES SILENT READING TEST [22]

TYPE 1. WORD RECOGNITION. Grades 1 and 2

FORM 1

did egg
dog two

be bed
bag she

may make
come milk

horse play
hose house

TYPE 2. WORD, PHRASE AND SENTENCE READING

1. A bed.

2. The dog is running.

TYPE 3. READING OF DIRECTIONS

1. Put an X on the ball.

3. Draw a line under the little book.

2. Put an X on the milk bottle.

4. Draw a line from the pig to the tree.

(26 Exercises)

[22] Copyright, 1926, by Bureau of Publications, Teachers College, Columbia University, New York.

Type A. Reading to Appreciate the General Significance

This is to be a reading test. You are to read a number of paragraphs. Below each paragraph are five words. One of the words tells how some one described in the paragraph felt—whether sad or happy, etc. You should draw a line under that one—and only one—word to show that you understand just how the person described in the paragraph did feel. Now let us try a sample before we begin the real test. Read the following paragraph and then draw a line under the word which you think tells best how the person felt.

> Once upon a time a young fairy went down to the river to swim. She jumped in with a splash. She put out her hands and tried hard to swim. Something seemed to be dragging her down. Oh, it was her wings! She had forgotten to take them off. Fairy wings become heavy when they are wet. She cried for help as loudly as she could.
>
> Draw a line under the word which tells how the fairy felt.
>
> cross angry weary afraid joyful

(16 exercises)

Type B. Reading to Predict the Outcome of Given Events

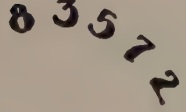
This is to be a reading test. You are to read a number of paragraphs. Below each paragraph are four sentences. Each sentence tells what is most likely to follow after the happenings that are described in the paragraph. You should draw a line under one—and only one—of these sentences to show that you can tell what will probably happen next. Now let us try a sample before we begin the real test. Read this paragraph and then draw a line under the one sentence which you think tells what will happen next.

The grocery man had a black cat. He loved his cat very much. One day a lady brought a big bulldog into the store. The grocer's cat raised his back and said "Meow! Psst!" to the bulldog. Of course, the dog did not like that, so he growled loudly. Before the grocery man or the lady knew what was happening, the bulldog had sprung upon the cat.

They let the fight go on
The cat slept on
The lady took her bird away
The grocery man saved his cat

(24 exercises)

TYPE C. READING TO UNDERSTAND PRECISE DIRECTIONS



3. The third grade boys have a baseball team. It is made up of the best nine players in that grade. They have played seven games this spring and have won five. They hope to win all the others. Draw a line under the number that tells how many games they have played.

7. This miner is getting ready to go down into a gold mine. He has on a miner's hat with a torch on the front of it. He lights it just before he goes into the dark tunnels. Make a cross where the miner will carry his light.

(24 exercises)

Type D. Reading to Note Details

This is a reading test. You are to read a number of paragraphs. Below each paragraph are three questions which you must answer by drawing a line under the one word or phrase which gives the best answer. Let us try a sample before we begin the real test. First read the paragraph. Then underline one—and only one—of the four answers to each of the questions to show that you understand what the paragraph said.

> Next morning she awoke and found herself in a beautiful room. The walls were covered with silken curtains. There were two mirrors made of pure silver. The bed was made of ivory. The coverings were made of silk and velvet. By her bed lay a dress and a pair of slippers. The dress was made of silk. The slippers were covered with diamonds.
>
> Where did the girl find herself?
> barn room garden store
> What were the mirrors made of?
> silver gold pearl silk
> What were on the slippers?
> rubies pearls opals diamonds

(18 exercises)

The *Sangren-Woody Reading Test* is a very recent development in the field of the analytical type of reading test designed for elementary school use. It measures seven specific aspects of silent reading ability; viz., word meaning, rate, fact material, total meaning, central thought, following directions, and organization. Illustrations of the types of exercises used in Part 1, Word Meaning, Part III, Fact Material, and Part V, Central Thought are reproduced here on pages 309–311.

Specimen Exercises from Sangren-Woody Reading Test [23]

PART I. WORD MEANING

Directions: Underline the word that makes the sentence true. Then write the number of that word in the parentheses.

Sample: A book is something to (1) wear, (2) read, (3) eat, (4) sail.....(2)

Begin here.

1. A dress is something to (1) **drink,** (2) **eat,** (3) **wear,** (4) **throw**.....( )
2. A letter comes in the (1) **water,** (2) **wind,** (3) **store,** (4) **mail**.....( )
3. A rock is a big (1) **stone,** (2) **animal,** (3) **train,** (4) **flower**.....( )
4. To buy something means to (1) **drink,** (2) **pay for,** (3) **run,** (4) **play**.....( )
5. A yardstick is three feet (1) **heavy,** (2) **hard,** (3) **long,** (4) **smooth**.....( )

6. A lake is made of (1) **sand,** (2) **water,** (3) **hills,** (4) **oceans**.....( )
7. Spain is the name of a (1) **king,** (2) **country,** (3) **woman,** (4) **car**.....( )
8. An automobile is a kind of (1) **paper,** (2) **chair,** (3) **car,** (4) **horse**.....( )
9. A kettle is used for (1) **colds,** (2) **games,** (3) **fishing,** (4) **cooking**.....( )
10. Honest means to be (1) **funny,** (2) **old,** (3) **truthful,** (4) **pretty**.....( )

(Forty Exercises)

---

[23] Copyright 1927, by World Book Company, publishers, Yonkers-on-Hudson, New York.

PART III. FACT MATERIAL

Directions: Write the answer to each question on the dotted line. Use one word if possible.

The "lead" in your pencil is not made of lead. Long ago people had lead in their pencils; that is probably why the pencils we use are called lead pencils. Another mineral called "graphite" is now used. This mineral is taken from mines in the same way as coal or iron ore.

1. What did people use in their pencils long ago?.....
2. What mineral is used in the pencils now?.....
3. From what is the mineral graphite taken?.....

When people began to use graphite for pencils, they took it just as it came from the mines, sawed it into strips and put it into wooden casings. The lines these pencils made were too black, and the "lead" was so soft that it wore away very fast. Men afterward learned to mix other materials with the graphite to make the mark less black and the "lead" harder and more lasting.

4. Did people ever use graphite alone in the pencils?.....

(Twenty Exercises)

PART V. CENTRAL THOUGHT

Directions: Draw a circle around the letter before the statement that best tells the important thought of the paragraph.

---

1. Cod fishing is dangerous business. Thick clouds of fog often settle over the sea more quietly and quickly than showers of rain come down on us. In the thick fog, where one cannot see fifty feet, the men in the dories sometimes lose their way. Sometimes great storms come up and upset their little boats.

a. Cod fishing is very dangerous business.
b. Heavy fogs often settle over the sea.
c. Men cannot see fifty feet in the heavy fog.
d. Storms often upset the little boats.

2. There is not room enough in France for every one to have a large farm. The whole of France is in size about equal to Missouri, Kansas, and Nebraska; but it has seven times as many people as these three states. Each farm in France must be small, and its owner must cultivate it well to make it provide a living for the family.

a. France is about the size of Missouri.
b. Farmers in France raise many grapes.
c. Every one has a large farm in France.
d. In France the farms are very small.

(Ten Exercises)

It is, of course, impossible in the limited amount of space allotted to the subject of reading in this volume to give a complete evaluation of the various instruments available for diagnosis in this subject. The teacher or student interested particularly in the problems of reading measurement and instruction should study the list of reading tests given in the *Appendix* and secure as many of these as possible for study and analysis. In this way many additional aspects of the diagnosis of the problem of reading instruction in the elementary school may be revealed.[24]

152. Diagnosis of Work-Study Type of Reading; High School and College Level. Observation of the rate and quality of reading of the work-study type done by high school and college students gives plenty of evidence that the need for reliable and valid diagnostic and remedial work in reading is not confined to the elementary school grades. One of the recent contributions to this type of diagnosis in silent reading at the high school or college level is the *Iowa High School Reading Tests*[25] which have been mentioned in another connection on page 74. These tests represent an effort to go considerably beyond the mere general survey of two or three phases of silent reading ability. They are designed to cover a wide range of skills indispensable to effective reading of the work-study type in Grades 7 to 12, and above. The skills contributing to the student's ability to read and to work with books, which are measured by these tests are given on page 74 of this book.

[24] There is a possible danger here that the reader, because of the emphasis which is placed on the measurement and diagnosis of reading, in this discussion, may forget that the tests are not ends in themselves but are only means to an end. Whenever the teacher is studying tests he will do well to keep in mind this question: How can this specific test aid me in following the learning problems of my pupils clear through to the use of remedial material if needed?

[25] Jorgensen, A. N., and Greene, H. A., *Iowa High School Silent Reading Tests.* Bureau of Educational Research and Service, University of Iowa, Iowa City, 1927.

One of the most interesting and valuable aspects of the reading skills measured by the *Iowa High School Reading Tests* is the ability to use the index for the purpose of locating information. This test, which is Part A of Test 6 of the test booklet, is reproduced here in order to give a more exact idea of how this type of ability is tested. It will also give to many teachers a number of suggestions for the development of similar material for instructional purposes.

These *Iowa High School Silent Reading Tests* have been found to be very useful in connection with the analysis of high school and college students' difficulties in the so-called content subjects. Students scoring in the upper percentiles for their groups can almost be guaranteed to do successful work in such subjects as geography, history, etc., if their ability is really applied to the task at hand. Furthermore, many high school teachers of English are finding that a few weeks of "hospital" work in silent reading will significantly improve the accomplishment of their pupils in English as well as in other subjects. For this purpose valid and reliable diagnostic measures of silent reading ability at a high school level are an essential part of the modern teacher's equipment.

153. Remedial Drills for Work-Study Types of Reading Difficulties. In the development of remedial exercises in silent reading the demands of social utility seem to place the emphasis upon reading of the work-study type. An examination of the outline of reading skills presented on pages 289 to 292 makes this immediately apparent. Possibly one reason for the rather marked emphasis on the work-study type of reading to the exclusion of reading of the leisure types is that there is good reason to expect a considerable carry-over of skills from the work-study type of reading to the other type, due to the large number of common skills involved. For example, skill in recognition of word meaning which functions in work-study reading is probably similarly effective when the individual is reading solely for pleasure.

TEST 6. USE OF THE INDEX [26]

PART A. (Time allowance: 2 minutes)

Directions to the Pupil: This is a sample index. It is needed in connection with the following exercises on the use of the index.

INDEX

Alaska; Agricultural possibilities, 213, 214; Commerce, 214, 215; Exports, 214 (Fig. 147); Fisheries and forests, 210-212; Fur farms, 210; Imports, 214; Map (Fig. 129), facing 197.

Citizens; and citizenship, 23–30; defined, 24; boys and girls, 26; responsibility of, 29, 84–85; duties of, 29–30, 49–59, see Americans.

Communities; definition, see community; kinds of, 9–18; origin of, 11–12; growth of, 12–20; large and small, 15–18; cooperation of, 19–20, 42, 49, 113; and health, 31–43; and the protection of life and property, 45; and education, 65–66; and beauty, 72–85; money for expenses, 87–98; organization of, 99–111; dependent upon each other, 139.

Corn; in America, 87; Argentina, 282; United States exports (Fig. 190), 282–284; plant, 87; industrial uses of, 189.

Courts; city, 108–110; county, 125; state, 120–267; national, 260–268.

Dairy Products; 82–84; in northern states, 82–83; states leading in production, 83; see also cattle.

Denmark; commerce, 411, 412, exports, 80, 411 (Fig. 262), 411; Map (Fig. 258), 408; possessions, 242, 248; resources, 81–84.

Dyes; 148–155, 226–228; ancient, 148–150; aniline, 153–155; cochineal, 150–152; discovery of native American, 150; dye woods, see forest; manufacture in the U. S., 155; recent increase in value (chart 15), 154; substitutes for natural, 153, 226; value of aniline, 154–155 (Fig. 20), 158; see also, cochineal.

Flour; see Wheat.

Gas; 46–52; as fuel, 46; waste, 49–50; see also power.

Indiana; coal, 145; corn, 44; hogs, 88; oil, 159; wheat, 85.

Nervousness; 85; and school life, 108; physical defects, 110; preventable, 111; causes of, 112; habit, 111, 113; from tobacco, 363.

Tobacco; instruction at school, 3; economic injuries of, 201; forbidden to employees, 210; evils of nicotinism, 363–386.

Wheat; fertilizers for, 48; insects injurious to, 51; kinds of, 43; preparation of seed bed, 45; when to sow, 47; world's bread grain, 42; wheat diseases, 49.

Study the following samples:

1. Where (on what page as shown in the above index) will you find information on coal mining in Indiana? 145

2. Can you find information regarding oil in Indiana? ..... *Yes*

Answer the remaining exercises in a similar manner.

---

1. Next to what page can you find a map of Alaska? .....
2. On what pages does the longest continued discussion occur about corn? .....
3. What is the number of the figure which shows something about the export of corn by the United States? .....
4. Under what topic can you find additional references to gas? .....
5. On what page can a definition be found for citizenship? .....
6. What word would you look for to find additional information about citizenship? .....
7. On how many pages will be found information telling you about the possessions of Denmark? .....
8. Does the index tell you on what page you can find something about flour? .....
9. On what page can you find a figure about the exports of Denmark? .....
10. On how many pages is the subject of dyes discussed? .....

Score = number right ..... *x 2* = _____

(*Total possible score* = 20 *points*)

---

[26] From *Iowa High School Silent Reading Test*, Copyright, 1927, by A. N. Jorgensen and H. A. Greene. Published by Bureau of Educational Research and Service, Iowa City, Iowa.

A few illustrations of specific types of remedial exercises suited for use in silent reading of the work-study types are presented in the following pages in the hope that they may serve as a guide to teachers interested in the development of material of this type. Only a few samples from each field can be furnished.

Word Recognition. Exercises designed to develop skill in the recognition of new word meanings appear in a great many forms, as for example: (1) simple sentence completion; (2) agent-action; (3) action-agent; (4) action-effect; (5) effect-action; (6) identification; (7) opposites; (8) similars; (9) description; (10) phrasing; etc. The following "action-agent" exercises are taken from the *Horn-Shields Silent Reading Flash Card Exercises.*[27]

Agent-Action Exercises

Each of these cards tells the name of something. As I show the name of each thing, tell what it can do. For example, if I show the word baby, you might say, cry, or laugh, or kick. (Show here a series of prepared cards carrying suitable words.)

Matching Exercises [28]

The following exercise is a specimen of a matching exercise which may be used when skill in the recognition of word meanings has reached a slightly higher level.

The elephant	has a trunk
Cows	have horns
The squirrel	has a bushy tail
The pig	has a curly tail
The rabbit	has pink eyes
The swan (or giraffe)	has a long neck
The sheep	gives us wool

[27] Horn, E., and Shields, Grace, "Manual for Using the Horn-Shields Silent Reading Flash Card Exercises." Ginn & Company, Boston, 1923.

[28] Horn, E., and Shields, Grace, "Learn to Study Readers, Book One," Page 21. Ginn and Company, 1924.

As a sample of exercises suitable for drill on differentiating similar words, material such as the following is suggested in the Manual of Directions of the Horn-Shields "Learn to Study Readers" (page 17).

Exercises in Completing Simple Sentences

A lemon is	sorry	sore	sour	sweet
Ice is	called	could	cold	card
Birds build	most	next	nets	nests
Stars shine at	might	right	night	neat

Location of Information in Books. A significant factor in the child's use of reading for work-study purposes is his ability to locate information in books.[29] The following suggestions may prove helpful:

1. To develop in pupils an ability to use the index children should: (a) be taught the alphabet; (b) be drilled in arranging words in alphabetical order; (c) be drilled in finding answers to questions by use of the index; (d) be asked to prepare an index for books not provided with them.
2. To develop the ability to use a table of contents, pupils should: (a) be assigned lessons by topic or title; (b) find the assigned lessons in the text by means of the table of contents; (c) find additional sources of information on the assignment in the library.

As an illustration of materials designed for the development of these skills the following sample taken from the Horn-Snedaker-Goodykoontz, "Learn to Study Readers," Book V,[30] is presented.

[29] Excellent material on how to use the dictionary; ability to use the library card files; and ability to use keys, tables, graphs, etc.; and ability to use reference materials is presented in "Lessons on the Use of Books and Libraries," prepared by O. S. Rice and published by Rand, McNally, and Company.

[30] Horn, E., Snedaker, Mabel I., Goodykoontz, Bess," Learn to Study Readers," Book V. Ginn and Company, 1926.

Dictionary Test Lesson on Locating Words

To locate words rapidly in a dictionary you must be able to do two things well; first, understand the order in which words are arranged in the dictionary; and second, make use of the guide words at the top of each dictionary page which tells what words may be found on it. How well can you do these two things?

1. Arranging words in the order in which they would be found in the dictionary.

   Write the words of each line in their dictionary order. Do "a," then "b," and so on.

   a. stamp, check, knife, paper, frown
   b. basket, butter, biscuit, berries, bottle, byway
   c. party, pace, paint, page, paws
   d. circumspect, circumnavigate, circumference, circumstance

2. Using the guide words to help locate words. At the tops of four pages in one dictionary are the following guide words:

   Page 1. scorer.....scratch
   Page 2. scratcher.....sculk
   Page 3. scull.....sealer
   Page 4. sealer.....second

   The first word given is the first word on the page; the other word is the last word on that page. Can you tell what words would come between them? Where would "scorn" come? On which page would "scud" be found?

scud	seam	seal	season
scum	scour	script	seat
scrim	scowl	scrub	scout
seclude	scuttle	scythe	search
seaweed	sculpture	screw	scrunch
scrimp	scramble	seasoning	scourge

Organization of Material. The ability to organize what is read is a necessary part of the equipment of everyone who expects to become a good student. Organization of reading materials calls for a superior type of judgment. The fol-

lowing suggestions will aid teachers in developing a variety of types of practice in organization.

1. Practice in deciding upon the main thought of the paragraph or topic. Skill in reading to discover the main thought in a paragraph may be gained through exercises calling on the student to:
   a. State the main thought of a paragraph or of an entire topic or assignment in a sentence.
   b. Select the best statement of the main thought from a number suggested by the class.
   c. Prepare a single statement that tells the theme of a story or the part played by its leading characters.

2. Drill in outlining a story, an assignment, a reference reading, a poem, etc. Skill in outlining may be developed by calling on the student to:
   a. Determine the order in which events occur.
   b. Read an assignment to find exact parts or events which will fit into an outline previously presented.
   c. Select the central thought of the story or assignment.
   d. Read the assignment or story to select the points which may be placed under several main headings.
   e. List names of characters of the story, or list titles of pictures that might be made up from the story.
   f. Decide upon the main divisions of a story and list a number of points under each division.
   g. Prepare a topical outline of a story to be used as a guide in telling the story to others.
   h. Build an outline after hearing a story read orally.
   i. Tell in order the steps in some specific industrial process from the gathering of the raw material to the finished product.
   j. Prepare an outline upon a given problem or assignment where the material is to be collected from various sources.

3. Practice in Analyzing the Organization of Selections
   In a well-written selection each paragraph possesses a unity of its own. Several paragraphs will be closely related to each other, and often this relationship is indicated by appropriate subheading or subtitles. These subtitles when arranged make an outline. Authors not infrequently fail to observe the laws of unity either in the whole selection or within the paragraph. Sometimes two or more stories are

interwoven. Pupils can be trained to detect faulty paragraph construction or poor story organization by calling on them to:

a. Pick out the paragraphs of the story that are well organized; e.g., examine the paragraphs to see whether the author has a topic sentence in each and whether he has made each sentence in the paragraph related to the topic of the paragraph; check the sentences that are related and cross out those which are not related.
b. Pick out the paragraph that appropriately belongs under the title of the selection and draw lines through those that do not.
c. Select the parts vital to the central theme of the story.
d. Decide which parts should be omitted or changed to make the selection more interesting.

4. Practice in restating the substance of a difficult passage to convey the same idea in simplified form. Ability to restate the author's ideas in one's own words is important in comprehension and in organizing the material read. It can be improved by calling on the student to:

a. Restate the substance of a long paragraph in a sentence or two.
b. Condense a selection to approximately one-third, or one-fourth of its present length.

5. Practice in selecting the most appropriate title for a selection. The child who is able to supply an appropriate title to a paragraph or a selection just read shows not only that he understands what it is about, but that he has interpreted the various statements as parts of an organized unit. Tentative titles must be set up and tested out until finally one is secured which is broad enough to cover all that is said and yet specific enough to serve as an accurate cue to the contents. Exercises of this type probably yield as quick results in the form of increased organizing ability as any that can be suggested. To provide such practice call on the student to:

a. Suggest titles for selections of prose or poetry read silently.
b. Decide whether the author's title for a selection is well chosen and suggest a better one.
c. Compare titles of a series of selections as to their appropriateness in arousing and maintaining interest.

d. List descriptive titles for a selection, imagining the story a new one, and re-reading it under the title which is considered best.
e. Divide a story into parts and select titles for each.
f. Suggest subtitles that might aid in getting all the facts out of the selection.
g. Give titles to pictures which might be used in a motion picture scenario.
h. Write appropriate titles after listening to a paragraph read orally by a pupil.

154. Other Remedial Materials in Reading. It is impossible within the limited space of this volume to present illustrations of even the more significant types of remedial or instructional exercises which may be devised to parallel diagnosis in reading as now developed. Undoubtedly the further development of diagnostic instruments in silent reading will reveal many other aspects of the field which demand special instructional devices. Thus far diagnosis in reading confines itself mainly to such factors as rate, comprehension of word meaning, comprehension of sentence meaning, comprehension of paragraph meaning. Beyond this very little has been accomplished. The factors of organization and memory are commonly recognized as fundamentals of silent reading, but so far only slight progress has been made in identification of the skills underlying these factors or in the development and validation of materials whose function is to improve them.

Commercial materials designed for instructional and remedial use in reading have been on the market for many years. However, this material has mainly emphasized the instructional aspect of the problem for beginners in reading, and usually from the point of view of some particular method of teaching reading. At the present time there are several sets of material available, having for their purpose the development of silent reading skills of the work-study type. Among these are the *Horn-Shields Silent Reading Flash Card Exercises* to which reference has been made a

number of times in this text, and the *McCall-Crabbs Silent Reading Test Lessons.*[31] An excellent treatment of remedial and follow-up work in silent reading is also to be found in three bulletins prepared by Dr. Worth J. Osburn.[32] Other related material is listed in the *Appendix* of this book.

IV. Diagnosis and Remedy in Other Subjects

155. Diagnostic and Remedial Treatment of Other School Subjects. Obviously, the possibilities of developing remedial materials for use following diagnosis are not exhausted with the treatment of the fields of arithmetic, language, and reading. However, these are certainly the most important tool subjects in the curriculum from an educational and social point of view. As such they represent the fields in which diagnosis and remedial work is most likely to be in demand. Outside of the field of handwriting there are few others in which remedial instruction is necessary or can be effectively applied. Handwriting itself has been very adequately covered by the excellent treatment given it by Freeman.[33] Special devices have also been proposed by many authors of handwriting methods. Remedial work in spelling is largely limited to the development of a method of learning on the part of the pupil. From that point on the most significant single factor is the selection of a proper spelling vocabulary. The content subjects, such as history and geography, are themselves so largely limited by inadequate mastery of silent reading skills that the development of remedial work in these fields will probably await further mastery

[31] McCall, W. A., and Crabbs, Lelah Mae, "Standard Test Lessons in Reading." (94 lessons in separate books for each grade.) Bureau of Publications, Teachers College, Columbia University, New York.

[32] Osburn, W. J., "Remedial and Follow-up Work in Silent Reading." Public School Publishing Co., Bloomington, Illinois, 1925.

[33] Freeman, F. N., "Principles of Method in Handwriting." Eighteenth Yearbook of the National Society for the Study of Education, Part I, pages 60 to 73, 1919.

———"The Teaching of Handwriting." Houghton Mifflin Company, 1914.

of these basic skills. In the meantime the busy classroom teacher will probably find sufficient occupation and profit in the measurement, diagnosis, and remedial treatment of those fundamental subjects in which the goals of instruction are reasonably clean-cut.

V. Summary

This section concludes the third and final chapter discussing and illustrating the problems involved in the use of tests for the diagnosis of pupil difficulties and for the purpose of providing a basis for remedial instruction in three socially important school subjects. The discussions have necessarily been brief, but in their brevity they have been pointed for the purpose of emphasizing certain principles which are believed to be important. The illustrations have also been few and brief, but in these the student will possibly find suggestions for the development of material which in validity and in general efficiency will go far beyond the samples given.

From the discussion of the diagnostic and remedial possibilities in the field of arithmetic (Chapter IX) it is apparent that here is a subject in which some rather definite analytical and constructive work is being done. It is now possible to diagnose arithmetic disabilities with considerable accuracy in the fields of whole numbers, fractions, decimal fractions, percentage, denominate numbers, mensuration, interest and business forms, and problem solving. At the present time also effective remedial and corrective materials are available in whole numbers, common fractions, decimal fractions, and problem solving. The near future will undoubtedly see the rapid development of more effective instructional and corrective exercises in these as well as in other aspects of arithmetic.

The diagnostic and remedial program in language is decidedly less clean-cut than is the case in arithmetic. This is unquestionably due to a lack of definiteness in the iden-

tification of the fundamental language skills. Little has been done in the diagnosis of oral language disabilities, and but little more has been done in the development of corrective instruction in this field. Written language in certain forms may be measured with reasonable accuracy, but the causes lying at the foundation of difficulty in this field are still indifferently identified. Certain more specific language skills such as capitalization, punctuation, language usage, and certain grammatical situations can now be diagnosed with a fair degree of accuracy. However, the discouraging feature of this situation is that remedial instruction designed to improve these specific skills does not seem to carry over significantly into total success in the use of written language. That is to say, a child may be given well constructed drill on such things as punctuation, capitalization, language usage, etc., all of which, in theory at least, contribute to success in the use of written language without improving the general quality of his written language product. This means that one of two things is true, either our criterion of language success is inadequate, or the specific skills on which instructional emphasis is placed are not the proper ones. It must be apparent then that while the language problem is an old one, it is still with us. Much more work needs to be done in the development of more accurate analyses, more valid and reliable diagnostic instruments, and more effective corrective and instructional materials in the field of language.

The subject of reading has undergone some interesting and significant developments in recent years due largely to a more adequate realization of the social importance of certain types of reading skills. The fact that the main bulk of all reading done under present social conditions is of the work-study type is being recognized in the rapid shift in instructional emphasis from the oral to the silent reading field. This does not mean that there is not an important place for the development of oral reading skills, but it means that they will be developed in their own place and for certain purposes. Obviously these purposes are not in conflict

with those underlying the development of silent reading skills. The recognition of these specific functions of the various types of reading has made it possible for much progress to be made in the diagnosis of reading disabilities. Reading tests capable of furnishing results accurate enough for individual diagnosis are now available for such reading skills as word meaning, sentence meaning, paragraph meaning, sentence organization, paragraph organization, use of an index, alphabetizing and rate of reading. Considerable progress has also been made in the construction of drill materials intended to increase skill in these phases of reading ability. A number of illustrations of this type of material are given in this chapter.

A brief résumé of certain of the more significant problems treated in these eleven chapters of discussion of the use, interpretation, and application of the results of tests by the classroom teacher is given in Chapter XII, which is entitled "Tests and the Classroom Teacher."

SELECTED REFERENCES

Book, W. F., "Learning How to Study and Work Effectively." Ginn and Company, Boston, 1926.

Buswell, G. T., "An Experimental Study of the Eye-Voice Span in Reading." Supplementary Educational Monographs, No. 17, December, 1920, University of Chicago Press.

Buswell, G. T., "Fundamental Reading Habits; A Study of Their Development." Supplementary Educational Monographs, No. 21, 1922, University of Chicago Press.

Gates, Arthur I., "The Improvement of Reading." The Macmillan Company, New York, 1927.

Germane, C. E., and Germane, Edith G., "Silent Reading." Row, Peterson and Company, Chicago, 1922.

Gray, W. S., "Diagnostic and Remedial Steps in Reading." *Journal of Educational Research,* 4: 1–15; June, 1921.

Gray, W. S., "Remedial Cases in Reading; Their Diagnosis and Treatment." Supplementary Educational Monograph, No. 22, University of Chicago Press, 1922.

Gray, C. T., "Types of Reading Ability as Exhibited through Tests and Laboratory Experiments." Supplementary Educational Monographs, Vol. I, No. 5, August, 1917, University of Chicago Press.

Greene, H. A., "Measuring Comprehension of Content Material." Twentieth Yearbook of the National Society for the Study of Education, 1925; pp. 114–126; Public School Publishing Company, Bloomington, Illinois.

Horn, E., "A Constructive Program in Silent Reading." *Journal of Educational Research,* 3:336–344; May, 1921.

Horn, E., and McBroom, Maud, "A Survey of the Course of Study in Reading." University of Iowa Extension Bulletin No. 99, College of Education Series No. 3, February 1, 1924.

Horn, E., and Shields, Grace, "Flashcards for Silent Reading Drill." Ginn and Company, Boston, 1923.

Horn, E., and Shields, Grace, "Learn to Study Readers." *Manual of Directions.* Ginn and Company, Boston, 1924.

Horn, E.; Snedaker, Mabel, and Goodykoontz, Bess, "Learn to Study Readers," Book V. Ginn and Company, Boston, 1924.

Huey, E. B., "The Psychology and Pedagogy of Reading." The Macmillan Company, Chicago, 1908.

Judd, C. H., "Reading, Its Nature and Development." Education Monograph, No. 4, Vol. 2, University of Chicago Press.

Klapper, P., "Teaching Children to Read." D. Appleton and Company, Chicago, 1914.

O'Brien, J. A., "Reading; Its Psychology and Pedagogy." The Century Company, New York, 1926.

Osburn, W. J., "Remedial and Follow-up Work."

1. Silent Reading—vocabulary.
2. Silent Reading—the understanding of sentences.
3. Silent Reading—the understanding of paragraphs and total selection.

Public School Publishing Co., Bloomington, Illinois, 1925.

Thorndike, E. L., "Reading as Reasoning; A Study of Mistakes in Paragraph Reading." *Journal of Educational Psychology,* June, 1917, pp. 323–32.

Twenty-fourth Yearbook of the National Society for the Study of Education, Part I. Public School Publishing Company, Bloomington, Illinois.

Uhl, W. L., "The Use of the Results of Reading Tests as Bases for Planning Remedial Work." *Elementary School Journal,* 17:266–276, December, 1916.

Wiley, J. A., "Practice Exercises in Supervised Study and Assimilative Reading." Published by the author. Iowa State Teachers College, Cedar Falls, Iowa, 1922.

Yoakam, G. A., "Reading and Study." The Macmillan Company, New York, 1928.

CHAPTER XII

TESTS AND THE CLASSROOM TEACHER

I. Summary

This chapter is a brief summary of a few of the more outstanding problems involved in the use, interpretation, and application of the results of tests by the classroom teacher as presented in the foregoing chapters of this volume. Naturally such a summary is most effectively organized to parallel closely the treatment in the several chapters themselves.

156. Recognition of the Need for Tests in the Classroom. Teachers long have measured the results of their teaching efforts. However, it is only relatively recently that any large degree of accuracy has been injected into their methods of measurement. For many years the teacher's estimate was accepted as the sole measure of a pupil's ability or accomplishment. Studies of the reliability of such methods gradually began to cast a doubt on their accuracy. Accordingly, interested teachers and research workers began to seek for more dependable measures. This movement was fostered by the so-called "survey" movement among educators which appeared at about the same time. Possibly the survey movement itself was a product of the same spirit of unrest and dissatisfaction with educational methods which brought into being the measurement movement. The survey movement left in its wake a perfectly logical result—the establishment of many centers of interest in the possibilities

of the more exact evaluation of the results of educational practices. These later appeared as bureaus of educational measurements and research, many of which are still functioning. Very distinctive service has been rendered by these agencies through their work in the construction, standardization, and critical evaluation of educational measuring instruments.

The testing movement has passed through the first stages of its development. Ten years ago it was necessary to popularize the idea. Now the advantages of standardized and informal tests are recognized by most educators and by many laymen. Moreover, the tests themselves are being subjected by these same persons to careful analysis and study. The most enthusiastic students of educational tests are their own severest critics. The shortcomings of tests are coming to be realized. They are not mysterious instruments for the confusion of the uninitiated, but are useful devices for the assistance of the progressive educator. They are an exceedingly useful means to a most important end—the improvement of the conditions under which teachers teach and children learn. If they accomplish this they are justified.

157. Meaning of Educational Tests. The meaning of educational tests to the classroom teacher may be best made clear by a consideration of the characteristics which distinguish them from other types of measuring instruments in education. In the first place the educational test, which usually appears in a standardized or semi-standardized form, is more limited and specific in its purpose than the ordinary informal examination. That is to say it usually confines itself to some relatively narrow field of subject matter. The typical examination or test prepared by the classroom teacher frequently covers a wide range of information. In the second place the exercises comprising educational tests are commonly constructed and arranged in accordance with certain statistical and educational principles designed to produce more accurate

measuring instruments. In the third place the more useful educational tests are standardized, that is, have been used by a large sampling of school children under controlled conditions. From this use of tests, or this standardization the forms and standards which give meaning to test results and permit the interpretation of test scores are derived. In the fourth place, the more carefully constructed and valuable educational tests yield results which point the teacher's way to the application of specific remedial methods where needed.

Educational tests are characterized by other features such as validity, reliability, objectivity, etc. Validity refers to the truth of the picture of the ability or achievement revealed by the test. Reliability refers to the consistency with which the test reveals this picture. Objectivity refers to the extent to which the test results are affected by the personal judgment of the user. There are other important characteristics of educational tests but these represent the major ones on which their meaning depends.

158. Uses of Educational Tests. Even if educational tests have failed to accomplish all of the wonders credited to them by enthusiasts they do perform certain more limited tasks so well that they are justified without the support of extravagant claims. Briefly, educational tests because of their definiteness and objectivity reveal to the teacher the status of the achievement of her class. They point out individual differences in capacity and achievement. They set up specific goals of achievement for the teacher. They reveal the results of special types of emphasis, or of special methods of instruction. They open to the administrator and teacher hitherto untouched sources of information useful in giving the pupil proper educational and vocational guidance. In their modern conception they reveal to the teacher the specific weakness of individual pupils so definitely that the teacher is in a position to apply effective instructional and corrective methods. Tests themselves have little or no power to bring about changes in pupil achievement as a mere

result of their use. Their chief service is their power to reveal pupil strengths or weaknesses. The correction of weaknesses is another aspect of the supervisory problem.

159. Using Tests in the Classroom. There is a growing feeling on the part of teachers that they themselves must assume a larger share of the responsibility for the use of educational tests in the classroom, and for the interpretation and application of the results after the tests have been given. Only by so doing does the teacher receive an adequate return from the use of tests. If this responsibility is to be wisely assumed the teacher must have an appreciation of the possibilities and the weaknesses of tests. She must be trained in their use and interpretation. She must be willing to exchange a certain amount of personal effort for the information concerning her teaching problems which the tests can furnish her.

Training in the use of tests comes as a result of their use. Opportunity for this training may be afforded through the preparation and use of objective informal tests as substitutes for the traditional examination, or it may be provided by undertaking a study of some supervisory problems of importance to the teacher in which standardized tests are used.

Main Uses of Tests. Three main types of uses of educational tests are noted, each resulting in different points of view regarding the teacher's responsibility. Tests of a detailed diagnostic type designed to give the teacher precise information concerning the abilities and limitations of her pupils are instructional in their function. The responsibility for the use of such material should be the teacher's. Tests designed to be used more particularly for survey or supervisory purposes should probably be administered by persons other than the teacher. The use of tests for administrative purposes such as for pupil classification, gradation, or sectioning, may well be the joint responsibility of the teacher and the administrator. In other words, the function the

tests are intended to perform determines where the responsibility for their use and interpretation lies.

Selection of the Test to Use. The common criteria for tests, validity, reliability, objectivity, administrability, and economy, afford the teacher a tangible basis for the selection of the test to use for a certain purpose. In a general way, however, the teacher should depend upon the advice of persons who have made a special study of the tests, rather than to attempt to apply these criteria personally. Information on each of these important aspects of tests is furnished with most of the better and more recently developed tests. If affirmative answers to each of the following questions are available concerning a specific test the teacher may feel reasonably safe in selecting it for use.

1. Does this particular test measure the abilities, habits, skills, attitudes, and information which I wish to measure?
2. How much time does it take to give the test? Is it long enough to give a reliable and consistent measure? (An unreliable measure is worse than useless.)
3. Is it easily and accurately scored?
4. Has it been widely used elsewhere?
5. Does it furnish accurate and extensive norms and standards for comparison and interpretation?
6. Is the interpretation of the scores simple and clear?
7. Do the results point the way to a remedial program?
8. Is the test economical in terms of time and money cost per unit of reliable information furnished by it?

Administration of the Test. One of the distinctive features of standardized tests is that they must be given under conditions which closely approximate the conditions under which they were standardized if the results are to be meaningful. Accordingly the teacher should carefully follow the directions furnished with such tests. Time limits should be rigidly adhered to. Obviously, test scores may in many cases be increased by allowing a more liberal time allowance, but it should be remembered that the purpose of giving the test is not to see how high a score can be made, but to discover if possible a true measure of the abilities of the pupils tested.

The attitude and personality of the examiner is also an important item in the administration of a test. The whole purpose of the test is defeated if an unnatural response is obtained. In the giving of tests in the beginning elementary grades the greatest care must be exercised to secure the cheerful confidence of the pupils.

Scoring the Test Papers. Much of the value arising from the use of tests in the classroom is lost if the teacher herself does not score the papers. In the first place the act of scoring the papers forces her to study their content more carefully than she would otherwise do. Such careful scrutiny often reveals hitherto unsuspected features in the test. In the second place the teacher is brought into much closer contact with the pupil's special abilities and limitations by the act of personally scoring his test paper. For the majority of the better tests the task of scoring the papers is greatly simplified and objectified by the use of answer keys and scoring stencils.

The next step in the use of tests in the classroom is the summary and interpretation of the results.

160. Summarizing and Interpreting the Results of Testing. Skill in summarizing and interpreting test results is dependent upon the mastery of the following statistical techniques:

1. A knowledge of how to classify and tabulate data.
2. A knowledge of how to find the common measures of central tendency.
3. A knowledge of how to express the variability of data.
4. A knowledge of how to determine the relationship between two or more groups of data.
5. A knowledge of how to derive and use norms, standards, and derived scores for purposes of comparison and interpretation of test results.

In addition to the use the classroom teacher may make of these skills in the proper utilization of test scores there is the application she may make of them in the study of

current educational literature. Reports of progress in education are filled with the statistical terms and techniques listed above. The teacher can scarcely hope to keep abreast of the times in her profession if she is unable to read current educational literature understandingly.

161. After Testing, What? The topic *After Testing, What?* raises the question which is in the back of the mind of every classroom teacher and every supervisor who has used current standardized tests. Much of the early use of tests was futile due to the fact that such broad vague phases of educational achievement were tested that even though reliable results were obtained, nothing specific could be done about the situation. Furthermore a great deal of the early use of tests in the classroom was a matter of satisfying curiosity. Hundreds of thousands of dollars' worth of tests have been given to millions of children, taking thousands of hours of pupil and teacher time, with absolutely nothing ever coming of it. Teachers have a right to expect that something tangible will be given them in return for interruptions and for pupil time spent in testing. One way to insure this return is for the teachers themselves to take an active part in the program. A type of training, an attitude toward their profession, a clearer insight into the difficulties faced by their pupils, is thereby gained which may not come to them in any other way.

The results of supervisory tests given periodically for the purpose of checking the efficiency of the teacher's instruction should be revealed to the classroom teacher in terms of specific suggestions for the further improvement of the situation. Instructional and diagnostic tests used by the teacher in the classroom should furnish her with such specific information concerning the abilities and limitations of her pupils that she can begin at once a program of preventive and corrective instruction.

In spite of the fact that a great deal of attention is now being focused on the problems of developing corrective and

remedial materials to use for the follow-up program after testing, only a limited amount of such material is now available. It is noticeable also that the main bulk of such material is limited to two or three subjects, namely, arithmetic, language, and reading. Illustrations drawn from these three are given considerable attention in the discussion of the problems of diagnostic testing and remedial teaching.

162. Conclusion. Within the past score of years tests and measuring devices in nearly all subject matter fields have been developed. What the future of this movement will be no one can predict. In many ways the rapid development has been unfortunate for it has resulted in confusion on the part of the classroom teacher, the one who should profit most from the program. There is a danger of over-production without sufficient refinement. This is illustrated in the tendency to flood the market with more and more test devices, with only an indifferent and mostly inadequate attempt to answer the question of what to do after testing. The encouraging aspect of the situation appears in the fact that teachers themselves are becoming better informed in the technique of testing and are demanding that the corrective and remedial aspects of objective supervision shall materialize in the classroom. Unquestionably the greatest developments in measurements in the next decade will be in the refinement of existing measuring instruments to the point that accurate and reliable information concerning the individual difficulties of each pupil in each particular type of achievement may be obtained. The corollary of this development is, of course, the preparation of valid instructional and remedial materials for use when the weaknesses of pupils are revealed. When such material is available in the major subject matter fields the program of measurement in education will be justified and the instructional problems of the classroom teacher will be greatly simplified.

APPENDIX

In the various sections of this *Appendix* will be found material supplementing in many ways the discussions and explanations presented in the main body of the text:

A. Distributing Agencies for Tests
B. Classified List of Tests and Publishers
C. Suggested List of Remedial Materials
D. Illustrative Problems in Educational Measurements
E. Individual Pupil Test Record Form
F. Glossary

APPENDIX A

Principal Distributors and Publishers of Educational Test Material

This *Appendix* offers a representative list of distributors and publishers of educational test material selected mainly with a view to serving the different sections of the country. Undoubtedly a number of important publishers and distributors of test materials are omitted from this list. This, of course, is impossible to avoid. So many new agencies are undertaking the publication and distribution of tests and similar material that it is hopeless to attempt to make such a list all inclusive.

Bureau of Administrative Research, University of Cincinnati, Cincinnati, Ohio.
Bureau of Co-operative Research, University of Indiana, Bloomington, Indiana.
Bureau of Co-operative Research, University of Minnesota, Minneapolis, Minnesota.
Bureau of Educational Measurements and Standards, Kansas State Normal School, Emporia, Kansas.

Bureau of Educational Research, University of Oregon, Eugene, Oregon.
Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.
Courtis Standard Tests, 1807 E. Grand Boulevard, Detroit, Mich.
Eau Claire Book and Stationery Company, Eau Claire, Wisconsin.
Educational Test Bureau, Minneapolis, Minnesota.
Ginn and Company, Boston, Massachusetts.
The J. B. Lippincott Company, Philadelphia, Pennsylvania.
Public School Publishing Company, Bloomington, Illinois.
Rand, McNally and Company, Chicago, Illinois.
Research Service Company, Los Angeles, California.
Scott, Foresman and Company, Chicago, Illinois.
Teachers College Bureau of Publications, Columbia University, New York City, New York.
University of Chicago Press, Chicago, Illinois.
University Publishing Company, Lincoln, Nebraska.
World Book Company, Yonkers-on-Hudson, New York.

APPENDIX B

Classified List of Educational Tests and Publishers[1]

The list of educational tests presented in this *Appendix* is in no sense a complete catalogue of standard tests. It represents merely a suggested list of tests which have somewhat established their utility by meeting particular school needs. The list is arranged alphabetically and the authors accept no responsibility for the merit of any of these tests. A number of excellent bibliographies of tests provide more extensive lists. Among these bibliographies are the ones published by the World Book Company, New York, and by the Bureau of Education, Washington.

Arithmetic

1. *Buckingham Scale for Problems in Arithmetic.* Division 1, Grades III and IV; Division 2, Grades V and VI; Division 3, Grades VII and VIII. Forms 1 and 2. Public School Publishing Co., Bloomington, Ill.
2. *Cleveland-Survey Tests in Arithmetic.* Grades III to VIII. Public School Publishing Co., Bloomington, Ill.
3. *Compass Diagnostic Tests:*[2]

Test I: Addition of Whole Numbers; Grades **2, 3,** 4, 5, 6, 7, 8
Test II: Subtraction of Whole Numbers; Grades **2, 3,** 4, 5, 6, 7, 8
Test III: Multiplication of Whole Numbers; Grades **3,** 4, 5, 6, 7, 8
Test IV: Division of Whole Numbers; Grades **4,** 5, 6, 7, 8
Test V: Addition of Fractions and Mixed Numbers; Grades **5,** 6, 7, 8
Test VI: Subtraction of Fractions and Mixed Numbers; Grades **5,** 6, 7, 8

[1] A corresponding list of Drill and Remedial Materials appears on pages 353 to 355 of this *Appendix*.

[2] Grade numbers in bold face indicate the grades in which the various tests are specifically intended for group use. They will be found more valuable for individual use in the remaining grades indicated.

Test VII: Multiplication of Fractions and Mixed Numbers; Grades 5, 6, 7, 8
Test VIII: Division of Fractions and Mixed Numbers; Grades 5, 6, 7, 8
Test IX: Addition, Subtraction, and Multiplication of Decimals; Grades 5, 6, 7, 8
Test X: Division of Decimals; Grades 6, 7, 8
Test XI: Addition and Subtraction of Denominate Numbers; Grades 6, 7, 8
Test XII: Multiplication and Division of Denominate Numbers; Grades 6, 7, 8
Test XIII: Mensuration; Grades 6, 7, 8
Test XIV: The Basic Facts of Percentage; Grades 6, 7, 8
Test XV: Interest and Business Forms; Grades 7, 8
Test XVI: Definitions, Rules, and Vocabulary of Arithmetic; Grades 5, 6, 7, 8.
Test XVII: Problem Analysis: Elementary; Grades 5, 6
Test XVIII: Problem Analysis: Advanced; Grades 7, 8
Test XIX: General Problem Scale: Elementary; Grades 5, 6
Test XX: General Problem Scale: Advanced; Grades 7, 8
Scott, Foresman and Co., Chicago, Ill. Also distributed by Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.

4. *Compass Survey Tests in Arithmetic.* Two forms, A and B, Elementary Examination for Grades II, III, and IV. Advanced Examination for Grades IV to VIII inclusive. Scott, Foresman and Company, Chicago. Also distributed by Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.
5. *Courtis Standard Research Tests, Series B.* Grades IV to VIII. Forms 1, 2, 3, and 4. S. A. Courtis, 1807 E. Grand Boulevard, Detroit, Mich.
6. *Knight-McClure Arithmetic Neatness Scale.* Scott, Foresman and Company, Chicago.
7. *Los Angeles Diagnostic Arithmetic Tests.* Grades II to VIII; Forms 1, 2, 3, 4. Research Service Company, 4259 South Van Buren Place, Los Angeles, Cal.
8. *Lunceford Diagnostic Tests in Addition.* Primary Grades; Forms 1 and 2. Bureau of Educational Measurements and Standards, Kansas State Normal School, Emporia, Kans.
9. *Monroe Diagnostic Tests in Arithmetic.* Part I, Integers, Grades IV to VIII; Part II, Integers, Grades IV to VIII; Part III, Common Fractions, Grades V to VIII; Part IV, Decimal

Fractions, Grades VI to VIII. Public School Publishing Company, Bloomington, Ill.

10. *Monroe General Survey Scales in Arithmetic.* Scale 1, Grades III, IV, and V; Scale 2, Grades VI, VII, and VIII; Forms 1, 2, and 3. Public School Publishing Co., Bloomington, Ill.
11. *Monroe Standardized Reasoning Tests in Arithmetic.* Test 1, Grades IV and V; Test 2, Grades VI and VII; Test 3, Grade VIII; Forms 1 and 2. Public School Publishing Co., Bloomington, Ill.
12. *Otis Reasoning Test in Arithmetic.* Grades IV to IX; Forms A and B. World Book Company, Yonkers-on-Hudson, New York, and Chicago, Ill.
13. *Peet-Dearborn Progress Tests in Arithmetic.* Intermediate Series, Grades IV and V; Upper-Grade Series, Grades VI, VII, and VIII. Houghton Mifflin Company, Chicago, Ill.
14. *Reavis and Breslich Diagnostic Tests in the Fundamental Operations of Arithmetic and in Problem Solving.* Grades VII, VIII, and IX. University of Chicago Press, Chicago.
15. *Spencer Diagnostic Test in Arithmetic.* Test 1, Grades III and IV; Test 2, Grades V and VI; Test 3, Grades VII and VIII. Bureau of Educational Research, University of Oregon, Eugene, Ore.
16. *Stanford Achievement Test.* See Stanford Achievement Test-Battery. World Book Company, Yonkers-on-Hudson, New York, and Chicago, Ill.
17. *Stevenson Problem Analysis Test.* Test 1, Grades IV to VI; Test 2, Grades VII to IX. Forms 1 and 2. Public School Publishing Co., Bloomington, Ill.
18. *Stone Reasoning Test (Revised).* Grades V to VIII. Bureau of Publications, Teachers College, Columbia University, New York City.
19. *Woody-McCall Mixed Fundamentals.* Grades II to VIII. Forms 1 and 2. Bureau of Publications, Teachers College, Columbia University, New York City.
20. *Wisconsin Inventory Tests.* Grades I to VIII. Houghton Mifflin Company, Chicago, Ill.

Batteries of Educational Tests

1. *Chapman's Classroom Products Survey Tests.* Grades V to VIII. Affords a means of measuring achievement in arithmetic and reading. J. B. Lippincott Company, 227 South Sixth Street, Philadelphia, Pa.

2. *Illinois Examination —Illinois General Intelligence Scale.* Monroe's Standardized Silent Reading Tests, Monroe's General Survey Scale in Arithmetic. Examination I, Grades III, IV, and V; Examination II, Grades VI, VII, and VIII. Public School Publishing Company, Bloomington, Ill.
3. *Indiana Composite Achievement Test.* Grade VIII. An extensive test battery covering elementary school achievement. Forms A and B. Bureau of Co-operative Research, Indiana University, Bloomington, Ind.
4. *Iowa High School Content Examinations.* Cover four fields of high school subjects: 1. English and Grammar; 2. Mathematics; 3. Science; 4. History and Social Science. Forms A, B, A-1, and B-1. Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.
5. *Lippincott-Chapman Classroom Products Survey Tests* (Arithmetic fundamentals, arithmetic problems, reading continuous passage, reading selections). Grades V to VIII. J. B. Lippincott Company, 227 South Sixth Street, Philadelphia, Pa.
6. *Otis Classification Test.* Tests both achievement and mental ability. Grades IV to IX. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
7. *Pintner Educational Survey Tests* (Arithmetic, reading completion, grammar, geography, and history). Grades II to VIII. Some of the sub-tests in this battery are abbreviated forms of well known tests. Pintner has also devised a battery of non-language mental tests which are designed to be used in connection with these educational survey tests. College Book Store, Columbus, Ohio.
8. *Pressey Second Grade Attainment Scale* (Reading, arithmetic, spelling). Forms 1 and 2. Public School Publishing Co., Bloomington, Ill.
9. *Pressey Third Grade Attainment Scale* (Spelling, arithmetic, and silent reading). Forms 1 and 2. Public School Publishing Co., Bloomington, Ill.
10. *Pressey Scale of Attainment No. 2* (History, arithmetic, and English). Grade VIII. Department of Psychology, Indiana University, Bloomington, Indiana.
11. *Stanford Achievement Test* (An extensive battery covering elementary school achievement). Primary Examination, Grades II and III; Advanced Examination, Grades IV to VIII. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.

Civics

1. *Brown-Woody Civics Test.* A test of vocabulary, information, and civic thinking. Form A. World Book Company, Yonkers-on-Hudson, N. Y.

English

Under the head of English are included measuring instruments for a variety of sub-divisions of the general field. Composition scales, tests in language, grammar, punctuation, and literature are listed here.

1. *Briggs' English Form Test.* Grades VII to XII. Forms Alpha and Beta. Bureau of Publications, Teachers College, Columbia University, New York City.
2. *Charters Diagnostic Language Test.* Pronouns, Verbs, Miscellaneous A and Miscellaneous B. Grades III to VIII. Forms 1 and 2. Public School Publishing Company, Bloomington, Ill.
3. *Charters Diagnostic Language and Grammar Tests.* Grades VII and VIII. Public School Publishing Co., Bloomington, Ill.
4. *Columbia Research Bureau English Test* (Steeves, Abbott, and Wood). For last years of high school work and college placement. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y.
5. *Franseen Diagnostic Tests in Language.* Range grades III to VIII. Forms A and B of 3 parts each. Part I, Pronouns; Part II, Verbs; Part III, Varied construction (exclusive of pronouns and verbs). Bureau of Administrative Research, University of Cincinnati, Cincinnati, Ohio.
6. *Hillegas Scale for Measurement of English Composition by Young People.* Grades IV to XII. Bureau of Publications, Teachers College, Columbia University, New York City.
7. *Hudelson English Composition Scale.*[3] Grades IV to XII. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
8. *Hudelson Typical Composition Ability Scale.* Grades IV to XII. Public School Publishing Co., Bloomington, Ill.
9. *Kirby Grammar Tests.* Grades VII to XII. Forms 1 and 2.

[3] The Hudelson Scale and the Nassau County Supplement to the Hillegas Scale are essentially revisions of this scale. In general they will be found more satisfactory than the original scale. The Thorndike-Extension of the Hillegas Scale is another revision which has corrected some of the faults of the original scale.

Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.

10. *Lewis English Composition Scales.* Grades V to XII. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
11. *Nassau County Supplement to the Hillegas Scale.*[4] Bureau of Publications, Teachers College, Columbia University, New York City.
12. *New York English Survey Tests.* Public School Publishing Company, Bloomington, Ill.
13. *Pressey Diagnostic Tests in English Composition* (Vocabulary, Grammar, and Punctuation). Revision of original tests now published by Public School Publishing Company, Bloomington, Ill.
14. *Starch Punctuation Scale.* Public School Publishing Co., Bloomington, Ill.
15. *Van Wagenen English Composition Scales.* Grades V to XII. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
16. *Van Wagenen Reading Scale for English Literature.* Forms A, B, and C. Public School Publishing Co., Bloomington, Ill.
17. *Wakefield English Tests* (for secondary school use). Forms A and B. Bureau of Administrative Research, University of Cincinnati, Cincinnati, Ohio.

Geography

1. *Buckingham-Stevenson Information Problems Test in United States Geography.* Grades VI to IX. Two forms. Public School Publishing Co., Bloomington, Ill.
2. *Courtis Supervisory Test in Geography.* Test A, Grades V-B to VI-A; Test B, Grades IV-A to VII-A. Forms A and B. S. A. Courtis, 1807 E. Grand Blvd., Detroit, Mich.
3. *Gregory and Spencer's Geography Test.* Grades VI to IX. Forms A, B, and C. Bureau of Educational Research, University of Oregon, Eugene, Ore.
4. *Hahn-Lackey Geography Scale.* Grades IV to VIII. H. H. Hahn, State Normal School, Wayne, Nebr.
5. *Indiana Geography Test.* Grades VII and VIII. Taken from the Indiana Composite Achievement Test. Bureau of Co-operative Research, Indiana University, Bloomington, Indiana.

Handwriting

1. *Ayres' Handwriting Scale* (Gettysburg Edition). Elementary or High School. Russell Sage Foundation, New York City.

[4] Ibid.

2. *Freeman Chart for Diagnosing Faults in Handwriting.* Range: All grades. Houghton Mifflin Company, Boston.
3. *Gray Standard Score Card for Measuring Writing.* Public School Publishing Co., Bloomington, Ill.
4. *Leamer's Handwriting Scale.* Public School Publishing Co., Bloomington, Ill.
5. *Thorndike Scale for Measuring Handwriting.* Grades II to VIII. Bureau of Publications, Teachers College, Columbia University, New York City.
6. *West's Chart for Diagnosing Elements of Handwriting.* Public School Publishing Co., Bloomington, Ill.

History

1. *Barr Diagnostic Test in American History.* Series A and B. High School use. Public School Publishing Co., Bloomington, Ill.
2. *Columbia Research Bureau American History Test* (Carman, Barrows, and Wood). For high schools and colleges. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y.
3. *Gregory Tests in American History.* Series A and B. High School use. Bureau of Educational Research, University of Oregon, Eugene, Ore.
4. *Hahn Scale for Measuring Ability of Children in History.* Grades VII and VIII. H. H. Hahn, State Normal School, Wayne, Nebr.
5. *Harlan Test of Information in American History.* Grades VII and VIII. Public School Publishing Co., Bloomington, Ill.
6. *Iowa High School Content Examinations.* For description see "Batteries of Educational Tests."
7. *Iowa Test of General Information in American History* (DeGraff, Ruch, Greene). Grades VII and VIII. Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.
8. *Stanford Achievement Test.* See "Batteries of Educational Tests."
9. *Vannest Diagnostic Test in Modern European History.* Secondary school use. Bureau of Co-operative Research, University of Indiana, Bloomington, Ind.
10. *Van Wagenen American History Scales.* Information Scale A and Scale B; Thought Scale A and Scale B; Character Judgment Scale A and Scale B. Bureau of Publications, Teachers College, Columbia University, New York City.

Intelligence Tests

Individual Intelligence Tests

For individual testing the Stanford Revision of the Binet-Simon Tests is the most widely used. It is published by Houghton Mifflin Company, Boston, New York, Chicago, and San Francisco. An extension of the Binet-Simon Scale by F. Kuhlman has also been published by Warwick and York, Baltimore, Md. The World Book Company, Yonkers-on-Hudson, N. Y., has recently published the Herring Revision of the Binet-Simon Tests. This revision has received some very favorable comment.

Group Intelligence Tests

1. *Army Examination, Alpha.* For use in high schools and colleges. Bureau of Educational Measurements and Standards, Kansas State Normal School, Emporia, Kans.
2. *Dearborn Group Test of Intelligence, Series I, Revised Edition.* Grades I to III. J. B. Lippincott Company, 227 South Sixth Street, Philadelphia, Pa.
3. *Dearborn Group Test of Intelligence, Series II, Revised Edition.* Grades IV to IX. This series of general intelligence tests consists of two parts—General Examination C and General Examination D. They are non-verbal in character. J. B. Lippincott Company, 227 South Sixth Street, Philadelphia, Pa.
4. *Detroit First Grade Intelligence Test.* World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
5. *Haggerty Intelligence Examinations.* Delta 1, Grades I to III; Delta II, Grades III to IX. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
6. *Illinois General Intelligence Scale.* Grades III to VIII. Public School Publishing Co., Bloomington, Ill.
7. *Myers Mental Measure.* Grades I to XII. Forms 1 and 2. Newson and Company, 73 Fifth Ave., New York City.
8. *National Intelligence Tests.* Scale A and Scale B. Grades III to VIII. Forms 1 and 2 of each scale. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
9. *Otis Group Intelligence Scale, Primary Examination.* Grades I to IV. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
10. *Otis Self-Administering Test of Mental Ability.* Intermediate Examination, Grades V to IX. Higher Examination, Grades IX to XII and for college freshmen. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.

11. *Otis Group Intelligence Scale, Advanced Examination.* Grades VII to XII. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
12. *Pressey Primary Classification Test.* Grades I and II. This is a revision of the original Pressey Primer Scale which has been widely used. Public School Publishing Co., Bloomington, Ill.
13. *Pressey Intermediate Classification Test.* Grades III to VI. Forms A and B. Public School Publishing Co., Bloomington, Ill.
14. *Pressey Senior Classification Test.* Grades VII and VIII. Forms A and B. Public School Publishing Co., Bloomington, Ill.
15. *Terman Group Test of Mental Ability.* Grades VII to XII. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.

Latin

1. *Henmon Latin Tests.* 5 vocabulary tests and 4 sentence construction tests. Grades IX to XII. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
2. *Inglis Latin Test: General Vocabulary.* Grades IX to XII. Forms A, B, C, D, E. Harvard University Press, Cambridge, Mass.
3. *Ullman-Kirby Latin Comprehension Test.* Grades IX to XII. Forms 1 and 2. Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.
4. *White Latin Test—Vocabulary and Sentence Test.* Grades IX to XII. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Illinois.

Modern Languages

1. *Iowa Placement Examinations; Foreign Language Aptitude.* High School and Colleges, Forms A and B. Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.
2. *Wilkins Prognostic Test in Modern Languages.* World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.

French

1. *American Council Alpha French Test* (Henmon and Coleman). In two parts and two forms. Part I is a vocabulary test. Part

II is a silent reading and composition test. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y.

2. *American Council Beta French Test* (Greenberg and Wood). An elementary test in three parts; vocabulary, comprehension, and grammar. World Book Company, Yonkers-on-Hudson, N. Y.
3. *American Council French Grammar Test* (Cheydleur). A multiple-choice type test as an alternate for the American Council Alpha and Beta French test. World Book Company, Yonkers-on-Hudson, N. Y.
4. *Columbia Research Bureau French Test* (Meras, Roth, and Wood). For high schools and colleges. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y.
5. *Handschin Comprehension and Grammar Test A: French.* First year French. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
6. *Handschin Silent Reading Test A and B: French.* First and Second year French. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
7. *Henmon French Tests.* Four duplicate forms of the test. Two parts (vocabulary and sentences). World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
8. *Iowa Placement Examinations; French Training.* High Schools and Colleges. Forms A and B. Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.

German

1. *American Council Alpha German Test* (Henmon, Morgan, Purin, Hinz). Similar in scope and purpose to American Council Alpha French Test. World Book Company, Yonkers-on-Hudson, N. Y.
2. *Columbia Research Bureau German Test* (Purin and Wood). For high schools and colleges. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y.

Spanish

1. *American Council Alpha Spanish Test* (Buchanan, Crawford, and Keniston). Similar in scope and purpose to American Council French Test. World Book Company, Yonkers-on-Hudson, N. Y.
2. *American Council Beta Spanish Test* (Calcott, Williams, and Wood). Similar in scope and purpose to American Council

Beta French Test. World Book Company, Yonkers-on-Hudson, N. Y.

3. *Columbia Research Bureau Spanish Test* (Calcott and Wood). For high schools and colleges. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y.
4. *Handschin Silent Reading Test A and B: Spanish.* First and Second year Spanish. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
5. *Iowa Placement Examinations; Spanish Training.* High Schools and Colleges. Forms A and B. Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.

Mathematics

Algebra

1. *Columbia Research Bureau Algebra Test* (Otis and Wood). For high school and college use. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y.
2. *Douglas Standard Diagnostic Tests for First Year Algebra.* Series A (for the fundamental operations); Series B (for additional processes). Bureau of Educational Research, University of Oregon, Eugene, Ore.
3. *Hotz Algebra Scales.* Test 1, Addition and Subtraction; Test 2, Multiplication and Division; Test 3, Equation and Formula; Test 4, Graphs; Test 5, Problems. Series A and B. The former is an abbreviated form of the latter. Bureau of Publications, Teachers College, Columbia University, New York City.
4. *Illinois Standardized Algebra Tests.* Public School Publishing Co., Bloomington, Ill.

Geometry

1. *Columbia Research Bureau Plane Geometry Test* (Hawkes and Wood). Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y.
2. *Lane-Greene Unit-Achievement Tests in Plane Geometry.* Series of six tests arranged by instructional units. Forms A and B. Ginn and Company, Boston.
3. *Minnick Geometry Test.* Test A, Drawing of Figures; Test B, Stating hypotheses and conclusions; Test C, Recalling facts and figures; Test D, Selecting and organizing facts to produce a proof. Public School Publishing Co., Bloomington, Ill.
4. *Perry Geometry Tests (Plane).* Lafayette Printing Company, Lafayette, Ind.

5. *Schorling-Sanford Plane Geometry Test.* Grade X or XI. Bureau of Publications, Teachers College, Columbia University, New York City.

GENERAL

1. *Rogers Test for Diagnosing Mathematical Ability.* Grade IX. Bureau of Publications, Teachers College, Columbia University, New York City.

MUSIC

1. *Hutchinson Music Test.* A test of ability in silent reading and recognition from musical scores. Grades VII to XII. Public School Publishing Co., Bloomington, Ill.
2. *Kwalwasser-Ruch Test of Musical Accomplishment.* Grades IV to XII. Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.
3. *Kwalwasser Test of Musical Information and Appreciation.* Grades IV to XII. Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.
4. *Torgerson-Fahnestock Music Tests.* Part A tests certain theoretical knowledge; Part B tests ear training. Grades IV to IX. Public School Publishing Co., Bloomington, Ill.

READING

1. *Burgess Picture Supplement Scale for Measuring Ability in Silent Reading.* Grades II to VIII. Forms 1, 2, 3, and 4. Division of Education, Russell Sage Foundation, New York City.
2. *Courtis Silent Reading Test No. 2.* Grades II to VI. Forms 1, 2, and 3. Courtis Standard Tests, 1807 E. Grand Blvd., Detroit, Mich.
3. *Gates Graded Word Pronunciation Test.* Primary Grades. Teachers College Bureau of Publications, Columbia University, New York City, N. Y.
4. *Gates Primary Reading Tests.* Type 1. Word Recognition, Grades I and II; Type 2. Word, Phrase and Sentence Reading, Grades I and II; Type 3. Reading of Directions, Grades I and II. Teachers College Bureau of Publications, Columbia University, New York City, N. Y.
5. *Gates Silent Reading Tests.* Type A. Reading to Appreciate the General Significance; Type B. Reading to Predict the Outcome of Given Events; Type C. Reading to Understand Precise

Directions; Type D. Reading to Note Details. Grades III to VIII. Teachers College Bureau of Publications, Columbia University, New York City, N. Y.

6. *Gray Standardized Oral Reading Paragraphs.* Grades I to VIII. Public School Publishing Co., Bloomington, Ill.
7. *Haggerty Achievement Examination in Reading, Sigma 1.* Grades I to III. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
8. *Haggerty Achievement Examination in Reading, Sigma 3.* Grades VI to XII. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
9. *Holley Sentence Vocabulary Scale.* Grades II to XII. Public School Publishing Co., Bloomington, Ill.
10. *Iowa High School Silent Reading Tests, Form A* (Jorgensen and Greene). Grades VII to XII. Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.
11. *Monroe Standardized Silent Reading Tests.* Test I, Grades III, IV, and V; Test II, Grades VI, VII, VIII; Test III, High School. Forms 1, 2, and 3. There is no Form 3 of Test III. Tests I and II have been revised (see below) and the use of the revised forms is recommended. Public School Publishing Co., Bloomington, Ill.
12. *Monroe Standardized Silent Reading Tests, Revised.* Test I, Grades III, IV, and V; Test II, Grades VI, VII, and VIII. Forms 1, 2, and 3. Public School Publishing Co., Bloomington, Ill.
13. *Sangren-Woody Silent Reading Test.* Grades IV to VIII. Two forms. World Book Company, Yonkers-on-Hudson, N. Y.
14. *Stanford Achievement Test.* See "Batteries of Educational Tests."
15. *Stone Narrative Reading Tests.* Grades III to IX. Public School Publishing Co., Bloomington, Ill.
16. *Thorndike Scale Alpha 2. For Measuring the Understanding of Sentences.* Part I, Grades III to V; Part II, Grades VI to XII. Bureau of Publications, Teachers College, Columbia University, New York City.
17. *Thorndike-McCall Silent Reading Scales.* Grades II to XII. Forms 1 to 10. Bureau of Publications, Teachers College, Columbia University, New York City.
18. *Thorndike Test of Word Knowledge.* Forms A, B, C, and D. Bureau of Publications, Teachers College, Columbia University, New York City.
19. *Thorndike Visual Vocabulary Scales.* Scale A-2, Series X and Series Y; Scale B, Series X and Series Y. Grades III to VIII.

Bureau of Publications, Teachers College, Columbia University, New York City.

20. *Van Wagenen Reading Scales.* Grades VII to XII. Public School Publishing Co., Bloomington, Ill.

Science

BIOLOGY

1. *Grier's Range of Information Tests in Biology, Physics, Zoology, and Botany.* N. M. Grier, Hollins College, Hollins, Virginia.
2. *Ruch-Cossmann Biology Test.* Grades IX to X. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.

BOTANY

1. *Caldwell's Science Tests* (Botany, Chemistry, Physics, and Zoology). Publisher—no information.
2. *Grier's Range of Information Tests.* See BIOLOGY.

CHEMISTRY

1. *Caldwell's Science Tests.* See BOTANY
2. *Powers General Chemistry Test.* Scales A and B. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
3. *Rivett's Comprehension Test in Chemistry.* Test 7, Form A, General Information; Test 8, Form B, Equation Writing; Test 9, Form A, Problems on Hydrogen and Oxygen. B. J. Rivett. Northwestern High School, Detroit, Mich.

GENERAL SCIENCE

1. *Caldwell's Science Tests.* See BOTANY.
2. *General Science Examinations, Series A* (Giles, Thomas, Schmidt). Series of 22 tests. Eau Claire Book and Stationery Co., Eau Claire, Wis.
3. *Ruch-Popenoe General Science Test.* Parts 1 and 2. World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.
4. *Van Wagenen Reading Scales* (General Science). Scale A. See READING.

PHYSICS

1. *Caldwell's Science Tests.* See BOTANY.
2. *Camp's Iowa Physics Test.* Scale 1. Mechanics; Scale 2. Heat; Scale 3. Electricity and Light. Public School Publishing Co., Bloomington, Ill.
3. *Columbia Research Bureau Physics Test* (Farwell and Wood). For use in high schools and colleges. Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y.
4. *Starch's Physics Tests.* University Co-operative Company, 208 State Street, Madison, Wis.
5. *Thurstone's Physics Test.* World Book Company, Yonkers-on-Hudson, N. Y., and Chicago, Ill.

PHYSIOLOGY

1. *Grier's Range of Information Tests.* See BIOLOGY.
2. *Diagnostic Physiology Test.* Grades VII and VIII. M. C. Olmstead, Clarkston, Wash.
3. *Indiana Physiology Test.* Grades VII and VIII. Taken from the Indiana Composite Achievement Test. Bureau of Co-operative Research, Indiana University, Bloomington, Ind.
4. *Otis Classification Tests.* See BATTERIES OF EDUCATIONAL TESTS.

ZOÖLOGY

1. *Caldwell's Science Tests.* See BOTANY.
2. *Grier's Range of Information Tests.* See BIOLOGY.

SPELLING

1. Ayres Spelling Scale for Measuring Ability in Spelling. Grades III to VIII. Division of Education, Russell Sage Foundation, New York City.
2. *Buckingham Extension of Ayres Spelling Scales.* Grades III to VIII and High School. Public School Publishing Co., Bloomington, Ill.
3. *Courtis Standard Dictation Tests.* Grades II to VIII. Forms A and B. S. A. Courtis, 1807 E. Grand Blvd., Detroit, Mich.
4. *Iowa Spelling Scales.* Public School Publishing Company, Bloomington, Ill.
5. *Iowa Dictation and Spelling Test.* Grades II to VIII; Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa.

6. *Monroe Timed Sentence Spelling Tests.* Test 1, Grades III and IV; Test 2, Grades V and VI; Test 3, Grades VII and VIII. Public School Publishing Co., Bloomington, Ill.
7. *Sixteen Spelling Scales* (Seven S Spelling Scales). High School use. Bureau of Publications, Teachers College, Columbia University, New York City.

Stenographic Proficiency

1. *Blackstone Stenographic Proficiency Tests; Typewriting.* Five Forms, A, B, C, D, E. World Book Company, Yonkers-on-Hudson, N. Y.

APPENDIX C

Suggested List of Drill and Remedial Materials

In this suggested list of drill and remedial materials it will be noted that a majority of those named are for instructional or drill purposes rather than for remedial purposes. Materials which in the opinions of the authors of this book are definitely designed for remedial use are so indicated. Due to the nature of certain elementary school and high school subjects, drill and remedial materials are entirely lacking.

ALGEBRA

1. *Practice Exercises in Algebra,* Gonelly, Huff, Studebaker (*Remedial*). Scott, Foresman and Company, Chicago, 1927.
2. *Work Book in Algebra,* Myers, G. C.; Thomas, Elizabeth; Persing, Knibe M. The Harter School Supply Company, Cleveland, Ohio, 1927.
3. *New Type Drill Exercises in Elementary Algebra,* Newmen, Elsie F. Charles E. Merrill Company, New York, 1927.
4. *Instructional Tests in Algebra,* Schorling, R.; Clark, J. R.; Lindell, S. A. World Book Co., Yonkers-on-Hudson, N. Y., 1927.
5. *Exercises and Tests in Algebra,* Smith, D. E.; Reeve, W. D.; and Morrs, E. L. Ginn and Company, Boston, Mass.

ARITHMETIC

1. *Arithmetic Work Books,* Ruch, G. M.; Knight, F. B.; Studebaker, J. W. Grades III to VIII. Scott, Foresman and Co., Chicago.
2. *Courtis Standard Practice Exercises in Arithmetic.* A series of 48 lessons on cards covering the operations with whole numbers in grades 4 to 8. World Book Company, Yonkers-on-Hudson, N. Y.
3. *Economy Remedial Exercises in Whole Numbers (Remedial).* Forty-one specially designed exercises on cards permitting con-

tinuous and repeated use. Standardized for use in grades 4 to 8. Scott, Foresman and Company, Chicago.

4. *Economy Exercises in Problem-Solving in Arithmetic (Remedial).* Fifty-four exercise cards providing extensive practice on the recognized skills involved in problem solving. Cards may be used repeatedly. Standardized for use in grades 5 to 8. Scott, Foresman and Company, Chicago.
5. *Fowlkes-Goff Practice Exercises in Arithmetic.* Designed for use in Grades 7, 8, and 9. The Macmillan Company, New York.
6. *Lennes' Work, Drill, and Test Sheets.* Laidlaw Brothers, Chicago.
7. *Studebaker's Practice Exercises in Arithmetic.* Scott, Foresman and Company, Chicago.
8. *Osburn's Corrective Exercises in the Fundamentals of Arithmetic (Remedial).* Houghton Mifflin Company, Boston, Mass.
9. *Thorndike Practice Books in Arithmetic.* Rand, McNally and Company, Chicago.

FRENCH

French Work Book, Adams, M., and Knight, F. B. Rand, McNally and Company, Chicago.

GEOMETRY

1. *Work Book in Intuitive Geometry,* Betts, Wm.; Miller, A. B., and Miller, F. B. The Harter School Supply Company, Cleveland, Ohio, 1927.
2. *Rapid Drill Exercises* (Flash Cards) *in Geometry,* Lane, R. O.; Knight, F. B., and Ruch, G. M., Scott, Foresman and Company, Chicago.

HANDWRITING

1. *Courtis Standard Practice Tests in Handwriting.* World Book Company, Yonkers-on-Hudson, N. Y.
2. *Leamer's Diagnostic Practice Sentences in Handwriting.* Grades II to VIII. Public School Publishing Co., Bloomington, Ill.

LANGUAGE

1. *Clark's Language in Use.* Teachers Manual. Public School Publishing Co., Bloomington, Ill.
2. *Wohlfarth's Self-Help Methods of Teaching English.* World Book Company, Yonkers-on-Hudson, N. Y.

READING

1. *Silent Reading,* Cram, Fred. D. Bulletin of Iowa State Teachers College, Cedar Falls, Iowa.
2. *Horn-Shields Flash Cards for Silent Reading Drill.* Ginn and Company, Boston, Mass.
3. *Osburn's Remedial and Follow-up Work in Reading (Remedial).* Public School Publishing Company, Bloomington, Ill.
4. *Wiley's Practice Exercises in Supervised Study and Assimilative Reading.* Published by the author, Prof. J. A. Wiley, Iowa State Teachers College, Cedar Falls, Iowa.

APPENDIX D

Problems in Educational Measurements

The following problems supplement the material presented in the earlier sections of this text and are added for the purpose of furnishing the interested reader who does not have access to the more extensive *Work-Book*[1] with opportunity for additional experience in the interpretation of various phases of test results. Problems illustrating a number of the more unusual and difficult situations have been carried to completion. In other cases only the data and the suggested form for the problem are presented, the organization of the data and the completion of the problem being left to the reader.

I. TABULATION OF TEST SCORES

Illustration No. 1. Arithmetic Test Scores of 21 Fourth Grade Pupils.

72, 71, 68, 25, 68, 43, 45, 66, 46, 66, 46, 63, 50, 58, 57, 52, 53, 55, 53, 55, 54.

The same scores *ranked* or arranged in descending order:

72, 71, 68, 68, 66, 66, 63, 58, 57, 55, 55, 54, 53, 53, 52, 50, 46, 46, 45, 43, 25.

The *range* of these scores is from 72 to 25, or a scale distance of 47 points. This range is useful in determining the number of *steps* or *class-intervals* to use in making a *frequency distribution.* (See Chapter VI, Summarizing the Results of Testing.) In making a frequency distribution, a sample of which is shown in Illustration No. 2, on page 357, it is not necessary to arrange scores in rank order prior to making the tabulation.

[1] Greene, H. A., "Work-Book in Educational Measurements," Longmans, Green & Co., New York, 1928.

Illustration No. 2. Arithmetic Test Scores of 21 Fourth Grade Pupils arranged in Frequency Distribution.

Class-interval	Mid-point	Scores	Frequency
70.5–73.5	72	//	2
68–70 [2]	69	//	2
65–67	66	//	2
62–64	63	/	1
59–61	60	0	0
56–58	57	//	2
53–55	54	////	5
50–52	51	//	2
47–49	48	0	0
44–46	45	///	3
41–43	42	/	1
38–40	39	0	0
35–37	36	0	0
32–34	33	0	0
29–31	30	0	0
26–28	27	0	0
22.5–25.5	24	/	1
Number of Steps = 17			$N = 21$

PROBLEM 1

Geography Test Scores of 21 Fifth Grade Pupils.

a. Arrange or rank these scores in descending order:

43, 15, 18, 43, 20, 43, 26, 42, 29, 40, 29, 38, 36, 36, 30, 35, 32, 35, 32, 32, 32.

b. What is the largest score made on this test by any pupil in the class?

c. What is the lowest score made on this test by any pupil in the class?

d. What is the range of these scores?

[2] These limits mean that 68's, 69's, and 70's are included in this interval.

e. If a frequency table of ten steps is decided upon how many units will there be per step?

f. What will be the limits of the class intervals used?

g. Tabulate these scores in a Frequency Distribution as explained in Chapter VI.[3] Choose your own size of step and number of class-intervals. Remember to choose steps with mid-points divisible by the size of the class-interval. Do all of your work on the left half of a sheet of paper and *preserve it for use in connection with other problems.*

PROBLEM 2

Reading Test Scores of 41 Fourth B Grade Pupils.

38, 36, 46, 71, 61, 49, 40, 67, 44, 52, 58, 51, 57, 21, 41, 32, 50, 80, 39, 77, 46, 57, 65, 58, 61, 72, 55, 57, 57, 61, 38, 66, 41, 66, 48, 33, 49, 36, 46, 34, 44.

a. Determine the range, set up a frequency table of a suitable number of steps, and tabulate the above scores. Do all of your work on the left half of a sheet of paper and *preserve it for further use* in connection with other problems appearing in this *Appendix.*

PROBLEM 3

Language Test Scores of 31 Sixth Grade Pupils.

50, 34, 34, 34, 45, 34, 33, 41, 33, 33, 40, 32, 32, 40, 32, 32, 37, 11, 14, 37, 37, 37, 29, 10, 31, 30, 24, 25, 29, 29, 29.

a. Determine the range, set up a frequency table with a suitable number of steps, and tabulate the above scores. Do all of your work on the left half of a sheet of paper and preserve it for further use in connection with other problems appearing in this *Appendix.*

[3] See particularly pages 124–126.

II. FINDING THE ARITHMETIC MEAN

Illustration No. 3. The Calculation of the Arithmetic Mean when no correction is necessary; the assumed mean and the true mean being identical.

Class Intervals	f	d	fd
40.5–43.5	4	+3	+12
38–40	2	+2	+ 4
35–37	4	+1	+ 4 (+20)
32–34	4	0	0
29–31	3	−1	− 3
26–28	1	−2	− 2
23–25	0	−3	0
20–22	1	−4	− 4
17–19	1	−5	− 5
13.5–16.5	1	−6	− 6 (−20)
	$N=21$		$\Sigma fd=0$

Total Number of Steps..... = 10
Class Interval..... = 3

Assumed Mean..... = 33
(Assumed mean happens in this case to be the true mean)

A. M..... = 33

Illustration No. 4. Computation of Arithmetic Mean from Frequency Distribution.

Class Intervals	f	d [4]	fd
70.5–73.5	2	+ 6	+12
68–70	2	+ 5	+10
65–67	2	+ 4	+ 8
62–64	1	+ 3	+ 3
59–61	0	+ 2	0
56–58	2	+ 1	+ 2(+35)
53–55	5	0	0
50–52	2	− 1	− 2
47–49	0	− 2	0
44–46	3	− 3	− 9
41–43	1	− 4	− 4
38–40	0	− 5	0
35–37	0	− 6	0
32–34	0	− 7	0
29–31	0	− 8	0
26–28	0	− 9	0
22.5–25.5	1	−10	−10(−25)
	$N=21$		Σfd [5] $=10$

$$\text{Correction} = \frac{\Sigma fd}{N} = +.476$$

$$= .476 \times 3 = +1.428$$

Class Interval = 3
Assumed mean = 54

A. M. (true mean) = the estimated mean plus [6] $\frac{\Sigma fd}{N} \times$ the number of units in the interval

A. M. = 54 + 1.428 = 55.428

[4] d = step deviations from assumed mean.
[5] Σfd = Algebraic sum of fd's.
[6] *Plus*, when sign is positive (+), *minus*, when sign is negative (−).

Problem 4

Computing the Arithmetic Mean from Frequency Distribution

a. Complete the steps in the computation of the arithmetic mean in the following frequency table.

Class Intervals	f	d	fd
167.5–172.5	1	+9	
165 [7]	1	+8	
160	3	+7	
155	3	+6	
150	11	+5	
145	22	+4	
140	30	+3	
135	45	+2	
130	78	+1	
125	86	0	0
120	56	−1	
115	48	−2	
110	23	−3	
105	6	−4	
97.5–102.5	5	−5	
	$N=416$		$\Sigma fd=$

Class Interval = 5
Assumed Mean = 125
Arithmetic Mean = the assumed mean plus the correction
Arithmetic Mean =

Answer (Arithmetic Mean) = 127.37

[7] Mid-points of intervals.

PROBLEM 5

Computing the Arithmetic Mean from Frequency Distribution

a. Complete the steps in the computation of the arithmetic mean in the following frequency table.

Class Intervals	f	d	fd
147.5–152.5	2		
143–147	3		
138–142	0		
133–137	0		
128–132	1		
123–127	1		
118–122	1		
113–117	2		
108–112	3		
103–107	2		
98–102	3		
93– 97	9		
88– 92	9		
83– 87	8		
78– 82	8		
73– 77	12	0	0
68–72	8		
63–67	6		
58–62	10		
53–57	8		
48–52	12		
43–47	8		
38–42	6		
32.5–37.5	3		
$N=$		$\Sigma fd=$	

Class Interval =

Assumed Mean =

Correction $= c = \frac{\Sigma fd}{N} =$

Arithmetic Mean = the Assumed Mean $+ \left(\frac{\Sigma fd}{N} \times 5\right) =$

Answer (Arithmetic Mean) = 74.88

Problem 6

Computing the Arithmetic Mean

Compute the arithmetic mean from the frequency table prepared for Problem 2. Any step in the table may be taken as the zero step or assumed mean without affecting the accuracy of computation of the mean. However, for practical purposes it is desirable to select a step approximating the probable location of the mean. In an ordinary distribution this point is often indicated by counting into the distribution until a half-way point is reached. The point of heaviest density of cases is also a partial guide in the selection of the step for the assumed mean. Answer = 50.85.

Problem 7

Find the Arithmetic Mean from the tabulation prepared for Problem 3. (Answer = 31.84 from frequency table with step of 5.)

III. FINDING THE MEDIAN

Illustration No. 5. Finding the Median or Mid-measure without tabulation.

The arithmetic test scores of 21 Fourth Grade Pupils used in Illustration No. 1 afford a suitable basis for demonstrating the *mid-measure* or so-called counting median, which is usually found by counting in toward the middle of test papers stacked in systematic order. The arithmetic scores as ranked in order are: 72, 71, 68, 68, 66, 66, 63, 58, 57, 55, 55, 54, 53, 53, 52, 50, 46, 46, 45, 43, 25. The mid-measure is the score of the middle paper or the average of the two scores nearest the middle when the number of cases is even. The mid-measure in this case is 55.

Problem 8

Find the mid-measure or counting median from the data presented in Problem 1. (Answer = 32.)

Problem 9

Find the mid-measure for the data in Problem 2. (Answer = 50.)

Problem 10

Find the mid-measure for the data in Problem 3. (Answer = 33.)

Illustration No. 6. Finding the Median from Frequency Distribution.

Class Intervals	*f*
167.5–172.5	1
165	1
160	3
155	3
150	11
145	22
140	30
135	45
130	78
125	86
120	56
115	48
110	23
105	6
97.5–102.5	3
	$N = 416$

The median score may be defined as the $\frac{N}{2}$ case in an array of measures. In order to find this case it is necessary to ascend the scale to a point where exactly one-half of the cases lie below it. In the accompanying frequency distribution the $\frac{N}{2}$ case (the 50th percentile) or the half-sum is 208. This 208th case falls some place in the step 122.5 to 127.5 of which 125 is the mid-point. ($3 + 6 + 23 + 48 + 56 = 136$ which is as near the half sum as it is possible to go without exceeding it). The proportion of this

step which must be taken is found by subtracting 136 (*the subtotal*) from 208 (*the half-sum*). This remainder, 72, becomes the numerator of the fraction of the step which must be taken. In this case the fraction is $\frac{72}{86}$ of the step which consists of five units. In decimal form this fraction equals 4.19 units which must be added to the beginning of the step, 122.5. The median therefore equals $122.5 + 4.19$, or 126.69.

Illustration No. 7. Determining the Median.

The median can be computed with equal accuracy by counting in from either end of the distribution. In fact, it is an excellent method of checking such computations for accuracy. In the accompanying distribution the half-sum $\frac{N}{2}$ is 62.5. Count down from the top of the distribution by adding $2+3+1+1+1+2+3+2+3+9+9+8+8=52$. Subtract 52 from the half-sum, 62.5, and 10.5 represents the number of cases which must be taken from the next step containing 12 cases. Thus, $\frac{10.5}{12} \times 5 =$ 4.375. Since this represents the distance we have gone into the step from the top the value 4.375 must be subtracted from the top limit of this step, which in this case is 77.5. The median thus becomes 73.125.

Class Intervals	f
147.5–152.5	2
145	3
140	0
135	0
130	1
125	1
120	1
115	2
110	3
105	2
100	3
95	9
90	9
85	8
80	8
75	12
70	8
65	6
60	10
55	8
50	12
45	8
40	6
32.5– 37.5	3
	$N=125$

77.5
−4.375

73.125 Class Median

Illustration No. 8. Determining the Median when it falls on both sides of a class interval of zero frequency.

In this distribution the half-sum falls in the step with zero frequency regardless of which way the count is made into the distribution of cases. In adding upward the half-sum takes all of the cases up to the lower limit of the step with a mid-point of 260. In adding downward all of the cases are taken to the upper limit of the same step. The best practice in this situation is to take the mid-point of the step itself as the median. Hence, in this case the most representative class median is 260.

RATE SCORES ON READING TEST

(Sixth Grade Pupils)

Words per Minute	Pupils
Class Intervals	*f*
372.5–387.5	0
365	2
350	0
335	1
320	2
305	0
290	3
275	2
260	0
245	5
230	2
215	1
200	0
185	1
162.5–177.5	1
	$N=20$

Illustration No. 9. Determining the Median when it falls at the junction point of two class intervals.

In this distribution, adding upward for the half-sum takes all of the cases to and through the step with a mid-point of 6. Adding downward takes all of the cases to the top limit of this same step. Since the range of scores tabulated as 6's is really 5.5 to 6.5, the median in this case is the top limit of the step, or 6.5.

NUMBER OF ARITHMETIC PROBLEMS SOLVED CORRECTLY

(20 4th Grade Pupils)

Class Intervals	*f*
11.5–12.5	0
11	2
10	1
9	2
8	1
7	4
6	2
5	3
4	3
3	1
2	0
1	1
	$N=20$

PROBLEM 11

Tabulating, finding Median, and the Arithmetic Mean

The following scores represent the number of arithmetic problems worked correctly by the individuals in a class of 29 pupils:

4, 11, 17, 12, 10, 9, 10, 11, 9, 11, 2, 14, 12, 10, 9, 14, 13, 16, 11, 13, 14, 11, 8, 8, 11, 9, 14, 12, 10.

a. Make a frequency distribution of these scores as shown in Illustration 1.

b. Compute the Median following the procedure shown in Illustrations 6 and 7. (Answer = 10.92.)

c. Find the Arithmetic Mean of the distribution. (Answer = 10.86.)

PROBLEM 12

Tabulating, finding Median, and the Arithmetic Mean

The following rate scores were derived from *Monroe's Standardized Sileent Reading Test,* Form 2, Test 2, Grade 7.

133	141	162	152	115	152	133	133	162
162	152	115	162	133	73	133	133	141
101	162	162	141	141	162	133	115	101
162	162	101	132	162	115	162	115	73
141	162	162	115	162	162			

a. Prepare a frequency table of these data.

b. Find the Median of the distribution. (Answer = 139.50.)

c. Find the Arithmetic Mean of the distribution. (Answer = 137.26, from frequency table with step of 5.)

IV. EXPRESSING THE VARIABILITY OF TEST SCORES

The range of scores, that is, the scale distance between the lowest and highest scores is one method of expressing variability. However, it is one of the least reliable measures since it is so easily affected by the variation of a single score. Range is found by merely deducting the smallest score in the array from the largest. In the case of the data in Illustration No. 1 the range is from 25 to 72, or 47 units. A much more reliable expression of variation is found in the "inter-quartile range" which is really the range of the middle 50 per cent. of the cases. In this way the operation of the unusual deviation of "sport" cases at the extremes of an array is eliminated. In actual practice this range of the middle half of the cases is expressed in terms of one-half of the inter-quartile range and is called Q. The third or upper one-fourth of the distribution is called Q_3 and the lower one-fourth is called

Q_1. The upper quartile and the lower quartile correspond respectively to the seventy-fifth percentile and the twenty-fifth percentile. The value Q and the median (Q_2) are distinct, however. One is a measure of *variation* or spread while the other is a measure of *central tendency.*

Illustration No. 10. Finding the Semi-inter-quartile Range (Q).

Assume for this illustration the data used in *Illustration No. 7.*

Step 1. Find Q_3 the third quartile (75th percentile) using the same method as is employed in determining the median, except that instead of counting into the distribution one-half of the cases you are to count in from the top end only one-fourth $\left(\frac{N}{4}\right)$ of the cases. In this illustration $Q_3 = 90.14$.

Step 2. Find Q_1 by counting in from the lower end of the distribution one-fourth of the cases also. The interpolation for exact values is exactly the same for quartiles as it is for the median. In this illustration $Q_1 = 53.91$.

Step 3. The semi-inter-quartile range or "Q" is found by subtracting the value of Q_1 from Q_3. The formula is $Q = \frac{Q_3 - Q_1}{2}$. In this illustration $Q = 18.11$.

PROBLEM 13

Find the semi-inter-quartile-range (Q) of the scores in the distributions prepared for (a) Problem 3, and (b) Problem 4.

(Answer: (a) 2.78; (b) 7.10.)

V. RELATIVE AND ABSOLUTE RANKS

Relative ranks are positions assigned to cases in terms of the array of which they are a part. The problem of relative rank was suggested in the data presented in Illustration No. 1. In this case the 20 scores were arranged in order of descending magnitude. The problem of ranking from this point on is merely that of assigning positions to the magnitudes as arranged. The largest score is given first position; the next largest, second; etc. The only difficulty involved appears in the case of two or more scores of the same size. The following illustration will make clear the method of handling this situation.

Illustration No. 11. Relative Ranking of Scores

Scores in order of magnitude	72,	71,	68,	68,	66,	66,	63,	58,	57,	55,
Positions assigned	1	2	3.5	3.5	5.5	5.5	7	8	9	10.5
Scores in order of magnitude	55,	53,	53,	53,	52,	52,	46,	45,	43,	25
Positions assigned	10.5	13	13	13	15.5	15.5	17	18	19	20

PROBLEM 14

Assignment of Relative Ranks

Rank and assign relative positions to the scores in Problems 1, 2, and 3.

PROBLEM 15

Computing Percentile Ranks

a. Following the method illustrated in Table 24 of Chapter VI of this text compute the percentile ranks to be assigned to the scores given in Problem 2.

b. Compute the percentile ranks of each of the pupils listed below on the two tests for which scores are given.

Pupil	Score		Pupil	Score		Pupil	Score	
	Test 1	Test 2		Test 1	Test 2		Test 1	Test 2
A	72	43	J	72	15	S	68	15
B	68	20	K	66	43	T	66	26
C	43	42	L	45	29	U	66	40
D	46	40	M	66	29	V	46	38
E	63	36	N	50	36	W	58	30
F	57	36	O	52	30	X	53	35
G	55	32	P	53	28	Y	55	25
H	71	42	Q	65	33	Z	49	18
I	45	21	R	60	34			

VI. USING TABLES OF NORMS

Problem 16

Finding Age Equivalents from Test Scores

Use the accompanying table of age equivalents for the *Stanford Achievement Tests* and complete the work begun in this list of pupil scores. The first two are filled in for illustration.

Age Equivalents of Composite Scores of Primary Examination

Composite Scores	Age Equivalent		Composite Scores	Age Equivalent		Pupil	Composite Scores	Age Equivalent	
	Yrs.	Mos.		Yrs.	Mos.			Yrs.	Mos.
6	7	6	25	10	3	1	9	8	0
7	7	8	26	10	4	2	14	8	9
8	7	10	27	10	6	3	10		
9	8	0	28	10	7	4	12		
10	8	2	29	10	9	5	19		
11	8	4	30	10	10	6	8		
12	8	6	31	11	0	7	17		
13	8	8	32	11	1	8	23		
14	8	9	33	11	3	9	18		
15	8	11	34	11	4	10	26		
16	9	0	35	11	6				
17	9	2	36	11	7				
18	9	4	37	11	9				
19	9	6	38	11	10				
20	9	7	39	12	0				
21	9	9	40	12	1				
22	9	10	41	12	3				
23	10	0	42	12	4				
24	10	1	43	12	6				

Problem 17

Comparing Test Scores with Grade Norms

Use the accompanying table of grade levels adapted from the norms for the *Haggerty Reading Examination Sigma 1* and complete the work begun in this list of pupil scores. The first two are filled in for illustration.

Table of Grade Levels

(Adapted from Norms for Haggerty Reading Examination Sigma 1)

Score	Grade	Score	Grade
4	1.0	12	2.0
5	1.1	13	2.2
6	1.2	14	2.5
7	1.4	15	2.7
8	1.5	16	3.0
9	1.6	17	3.2
10	1.7	18	3.5
11	1.8	19	3.8
		20	4.0

Pupil List

Pupil	Score	Grade Level
1	13	2.2
2	8	1.5
3	12	
4	19	
5	10	
6	16	
7	20	
8	14	
9	4	
10	15	

Problem 18

Finding Quotients from Ages Given

Complete the work begun in the accompanying list of pupils and scores. Use the accompanying table of age norms for the *Haggerty Intelligence Examination, Delta 1,* to determine age equivalents. The first two quotients have been correctly computed for illustration. Intelligence Quotients (I. Q.) are obtained by dividing mental age (expressed in months) as given in Column 4 by the chronological age (column 2) or life age of the child expressed in months. The decimal is multiplied by 100 to give a whole number.

Revised Age Norms for Intelligence Examination: Delta 1

Year	Month											
	0	1	2	3	4	5	6	7	8	9	10	11
6	30	32	34	36	38	40	42	44	46	48	50	52
7	54	55	56	57	57	58	59	59	60	61	62	63
8	64	65	66	66	67	68	68	69	70	70	71	71
9	72	73	74	74	75	76	76	77	78	78	79	79
10	80											

Pupil	Life Age		Point Score	Age Equivalent		Quotient
	Yrs.	Mos.		Yrs.	Mos.	
1	7	0	54	7	0	100
2	7	2	59	7	6	105
3	6	6	63			
4	8	5	58			
5	7	6	55			
6	7	0	36			
7	7	4	55			
8	8	1	42			

APPENDIX E

Cumulative Record Card for Tests

Name.....Z——— (Last) ..... Y——— (First) ..... Grade.....8.... Age..13..... Yr. .....6 Mo.

Individual Test Score Card

Mental Tests	1926				19—				19—				19—			
	Test Score	Test Age	Quotient	Grade Level	Test Score	Test Age	Quotient	Grade Level	Test Score	Test Age	Quotient	Grade Level	Test Score	Test Age	Quotient	Grade Level
Terman A.....	149	16–8	124	12+												
Achievement Tests																
Stanford A.....																
Reading.....	222	15–11	118	9+												
Arithmetic.....	280	17–6	130	9+												
Natural History	85	17–4	128	9+												
Hist. and Lit...	62	15–7	115	8.6												
Dictation.....	164	15–2	112	8.3												
Language.....	40	15–6	115	9+												
Composite.....	85.3	16–3	120	9+												

Remarks—Reverse Side of Card

APPENDIX F. GLOSSARY

For the convenience of the reader who may be unacquainted with many of the terms used in a somewhat special sense in this book this list of definitions is presented.

ability. The power to produce; the result of training operating on innate capacity.

accomplishment. Used in the sense of production, achievement.

accuracy. Ratio of number of exercises attempted to number correctly done.

***A.Q.*—the Accomplishment Quotient.** A ratio expressing relative educational development to mental development.

achievement. Used in the sense of accomplishment, production. Also used as descriptive of tests measuring abilities or products acquired primarily from school training.

administrability. One of the criteria for the selection of tests. It implies the ease with which the material may be used, scored, and interpreted.

age equivalent. The age score assigned to given point scores on tests provided with age norms.

age norms. The typical or expected achievement of subjects of various ages usually expressed as medians or as averages of actual performance of subjects of different age groups.

alternate response test. An objective test of the type which permits the subject to choose between two possible answers, one of which is right and the other of which is wrong. See also true-false tests.

analysis. Used in the sense of taking apart or reduction in the process of identifying specific skills.

aptitude test. Same as prognostic test.

arithmetic mean. The most commonly used of the measures of central tendency. Commonly called the average.

array. A collection of data usually organized around a particular function.

assumed mean. The mid-point of the class-interval which is taken as the point about which the correction is made in computing

the arithmetic mean by the short method from a frequency distribution.

average. A generic term covering the measures of central tendency, but commonly used to designate the arithmetic mean.

battery of tests. A group of several tests covering a number of different subject matter fields which are usually administered together. Illustrated by the Stanford Achievement Tests.

best answer test. A type of multiple response test.

capacity. Power to learn or to profit from training.

central tendency. A term corresponding to average commonly applied to the arithmetic mean, median, and the mode.

chronological age. The number of years the pupil has lived; his life age. May be expressed in either years or months.

class-interval. A limited series of consecutive points used as a pigeon-holing or grouping device in tabulation. Used somewhat synonymously with group interval or step.

classification. The process of assigning an otherwise unassigned pupil to the particular unit or grade in a school for which his abilities and training best fit him.

coefficient of correlation. An index of straight line relationship which ranges in value from plus one (+ 1.0) through zero to minus one (− 1.0). There are a number of indices of correlation, many of which are called coefficients. The use of the term coefficient of correlation has become somewhat restricted and refers generally to the Pearson Coefficient obtained by the so-called product-moment method.

comparable measures. Measures which are expressed in terms of the same unit or with respect to the same point of origin.

completion test. A common form of the new type of examination, consisting of statements in which one or more important words have been omitted. The test calls on the subject (individual tested) to supply the missing words.

composite score. A single value used to express the results obtained from the use of a number of different measures. This may be the mean of scores yielded by several tests after these scores have been changed into equivalent form.

comprehension. A score denoting the degree of understanding of an exercise or material read.

conditions. Factors operating to introduce variants into testing or experimental situations.

control group. The group which in an educational experiment is not subjected to the influence of the experimental factor used

error of grouping. A variable error introduced by the practice of placing together in class-intervals unlike measures in tabulation.

on the experimental group. The results from the two groups afford a basis for the evaluation of the experimental factor.

correction. A term used to imply a remedy or adjustment. Also used in a somewhat technical sense in the computation of the arithmetic mean, etc.

corrective. Having for its purpose the remedying of observed defects or difficulties.

correlation. The relation between two or more measures of the same individuals or items.

correlation table. A two-way or double entry table which shows the relation between pairs of measures of the same individuals or items.

criterion. The standard by which the validity of a test may be determined.

derived scores. Values obtained through the changing of form of certain test scores, as ages, quotients, percentiles, etc., for the purpose of making data from different tests and sources comparable.

deviation. The amount by which a score or other measure differs from the central tendency of such a series of measures. Thus, if a child has a score of 14 and the median is 10, his deviation is plus 4.

diagnosis. Exact identification and location of strengths or weaknesses.

difficulty. A quality which when present in a test exercise results in a large per cent of incorrect responses.

discrimination. The quality in a test which causes it to distinguish adequately between varying levels of ability.

dispersion. See *deviation.*

educational age. Somewhat synonymous with achievement age; however, it usually applies to a pupil's standing in a number of school subjects while achievement age frequently refers to a single subject.

***E.Q.*—Educational Quotient.** The ratio between the chronological age and the subject or educational age as found by a standard test. Found by dividing educational age as found by suitable educational tests by chronological age. Both ages are reduced to months and the resulting quotient is multiplied by 100 to keep it in whole number form. The quotient represents the relative educational development of the subject.

educational test. A term applied to a test which measures a pupil's mastery of subject matter taught in school.

equivalent form. A duplicate or equal form of a standard test which yields a score approximately equal to the original form.

error of sampling. A common difficulty arising in educational measurement resulting from the use of a limited number of cases taken as being typical of a larger group.

essay type examination. See *traditional examination.*

exercise. A unit of work in a test governed by a specific set of directions.

experiment. The process of controlling conditions so that a particular situation may be repeated with only one variable operating.

f. An abbreviation used in certain tabulations to mean frequency.

fd. An abbreviation standing for frequency-deviation, one of the steps used in finding the arithmetic mean from a frequency table. It is found by multiplying the step-deviation from the assumed zero by the number of cases found at that class-interval.

first quartile (Q). The point on a scale of values such that 25% of the measures are less and 75% are greater.

fore exercise. A preliminary or practice exercise for the purpose of giving the pupil experience with the particular test situation.

form. A term used to distinguish between two or more arrangements of similar but not identical test exercises. Usually two forms of a test are approximately equal in difficulty.

frequency. The number of measures in a given class interval of a frequency tabulation.

frequency curve. A graphic representation of a distribution of measures.

frequency table. A distribution showing the number of cases or measures at each class-interval when put in tabular form.

fulcrum. The axis upon which a lever is supported and rotated.

G-scores. Grade scores.

general intelligence test. A test designed to measure general native capacity. Contrasted with achievement test.

general survey test. A test designed to measure a number of different aspects of a single school subject or a test composed of a number of parts each covering a phase of a different school subject. Contrasted with diagnostic tests.

grade equivalent. The grade or portion of the grade nearest which a pupil's test score places him when compared with the norms of the test.

grouping. The process of classifying or tabulating data into certain steps or categories.

group test. A test designed to be given to a number of different individuals at the same time.

half-sum. A term used in connection with the finding of the median. It is equal to N/2.

individual differences. The observed or measured differences between individual pupils in capacity, ability, industry, etc.

informal test. A test prepared by the classroom teacher as distinguished from the standardized test.

***I.Q.*—the Intelligence Quotient.** An expression of relative brightness. Found by dividing mental age as obtained from an intelligence test by chronological age. Both ages are usually reduced to months and the quotient put into percentages.

intelligence. Defined most simply and practically as the power to learn or to profit from training.

interval. A limited series of scale points set up for grouping or pigeon-holing data in tabulation. Used somewhat interchangeably with "step."

interpolation. The process of locating an intermediate point between two known points in accordance with the laws operating in the case at hand.

interpretation. Used mainly in the sense of explaining results of test scores and applying them to concrete or specific situations.

mean. Usually signifies the arithmetic mean; the point on the scale about which the deviations are zero.

median. A common measure of central tendency. A point on the scale such that there are as many equal to or larger than as there are equal to or smaller than. The middle of a series of scores with these scores arranged in order of magnitude.

mental ability. Used somewhat synonymously with intelligence; power to learn.

mental age (*M.A.*). The mental ability of a subject expressed in terms of the age of an average child having that ability. Thus, if a child has a mental age of an average 10 year old child he is said to have a mental age of 10 years.

mid-measure. The middle measure of a series of values arranged in order of magnitude. Similar to the median.

mid-point. The exact middle of a step or class-interval in a frequency table. Used as the point about which to correct in finding the arithmetic mean by the shorter methods.

multiple choice test. A test composed of exercises arranged in such a way as to require the pupil to select one or more correct responses out of a group of several possible responses.

N. A symbol commonly used to designate the number of cases in an array.

negative correlation. A relationship in which the larger values of

one variable are accompanied by the smaller values of the other variable.

normal. Typical; making regular progress, growth, or development.

norms. The median or average performance of subjects of different ages or grades, as determined by the testing of large numbers of cases.

objective. A term used in describing tests in which no opportunity for disagreement as to correctness of response exists.

opposites test. A test consisting of a list of terms calling for the opposite meaning of each to be given.

percentile. The points which divide the total number of cases in a given frequency distribution into 100 equal parts.

performance. Accomplishment, achievement. Used also to distinguish test scores as such from capacity.

power tests. Devices which measure achievement in terms of the difficulty of the task which the subject is just able to perform.

practice effect. The effect which previous experience with a test has on a later trial.

prognostic test. A test having for its purpose the prediction of probable future achievement based upon present performance.

quality scales. Devices, such as the Ayres Handwriting Scales, which measure by comparison with a set of standard specimens the outcome of the application of some specific skill.

quartiles. The results of dividing a distribution into quarters. The first quartile is the lower one-fourth, the second quartile is the median, the third quartile is a point equalled or excelled by only one-fourth of the cases or measures.

random sample. The process of selecting a limited number of cases from an unlimited group without allowing personal bias to enter into the selection. A random sample is one selected on a purely chance basis.

range. The limits of the scale necessary to include all measures in a distribution. Found by taking the smallest measure in the array from the largest.

rank. Position assigned to a score in a series.

rate score. A score expressing a pupil's rate of work.

rate tests. Devices which measure achievement in terms of the number of tasks of uniform difficulty which can be performed in a specified time.

recall test. A new type of test exercise so stated as to be answered correctly by the use of a single word contrasted with recognition tests.

recognition test. A test containing exercises of the true-false or multiple-answer type. These exercises do not force the student to formulate the answer.

relative rank. Position assigned cases or scores in a limited array.

reliability. Applied in a somewhat technical sense to tests as a means of expressing the consistency of performance of the test itself.

remedial. Material and devices having for their purpose the correction and elimination of observed difficulties and weaknesses.

retarded. A term applied to an individual whose school progress, mental or educational development, is slower than is expected of the normal subject.

right-minus-wrong formula. Formula for correction of chance in scoring alternate-response exercises.

sampling. See *random sample.*

scale. This word is used in two different senses in measurement. In the narrow sense "scale" means the portion of every measuring instrument which is used to express the pupil's performance. Scale is also used in a general sense to designate a measuring instrument in which the scale for expressing performance is the distinguishing characteristic.

scores. A description of the performance of a subject.

sigma. The common abbreviation for standard deviation.

smoothed data. Data representing too few measures to be truly representative of a situation often show irregularity not typical of the entire group. The so-called process of smoothing, irons out these irregularities and approximates the type of distribution which would result if the sample were adequate.

standards. The level of achievement in a given subject to which a school should bring its pupils. Norms are usually medians or averages based on actual achievement of subjects while standards are goals to be attained.

standard deviation. One of the more common measures of deviation or variability of scores.

standardized tests. A test is called standardized when the exercises comprising it have been properly selected and evaluated, and when in addition norms or standards have been established for it.

standardization. The process of setting up suitable objective goals of performance for tests or other instructional materials.

step. A term used interchangeably with class-interval. Refers to interval in a frequency tabulation.

sub-total. A term used in connection with the finding of the median.

subjectivity. The degree to which measurement results are affected by personal factors or judgments. Objectivity and subjectivity are both relative terms. All tests are somewhat subjective; some are more so than others.

survey tests. Used to describe tests which have for the main purpose the measurement of abilities in terms of broad general functions.

tabulation. The process of grouping or classifying data for purposes of condensation and interpretation.

test. Used in a general sense to describe any type of instrument for measuring any mental ability. More specifically it is the portion of a measuring device by which the pupil's performance is secured.

traditional examination. Refers to the type of examination commonly used previous to the development of the so-called objective-type tests. In the main they consist of exercises of the discussion type.

true-false test. A form of recognition test in which the student is called upon to determine the truth or falsity of the exercises.

validity. A term used to express the degree to which a measuring instrument measures the thing it purports to measure, or the extent to which it measures a thing which the subject tested has had an opportunity to learn.

variable. Refers to a quality which may exist in different amounts.

variability. Synonymous with deviation.

yes-no tests. A variety of alternate response tests comparable to the true-false type.

INDEX

Abbot, A., 88
Absolute ranks, 169
Accomplishment quotient (*A.Q.*), 203–205
Achievement age, 188
Achievement tests, 48–58; 353 ff.
Administrability, 81–82
Age-norms, 188–191; interpretation of, 191–194
Algebra tests, 347; drills, 353
Alienation, coefficient of, 163
Analytical diagnosis, 41–44
Arithmetic, 231–250; diagnosis of, 233–239; remedial drills, 239-249; 353–354; skills, 233–239; tests, 337–339
Arithmetic mean, classroom use of, 132; definition of, 128; illustration of calculation of, 131; summary of steps in calculation of, 133
Array, 117
Asker, W., 114
Average, 127
Averill, L. A., 287
Ayres, L. P., 8
Ayres Handwriting Scales, 15

Ballenger, H. L., 259, 262
Barthelmess, H. M., 114
Basis for testing program, 70
Bibliographies, 8–9, 46, 65–66, 88, 114–115, 175, 205–206, 230, 250, 284, 325–326
Binet-Simon, 20
Blanton, Smiley, 257, 284
Book, W. F., 287, 325
Borden, R. C., 257
Briggs, T. H., 88, 252; form tests, 255
Brinkley, Sterling G., 92, 114
Buckingham, B. R., 202
Bureaus of research, 5
Burgess Reading Scales, 26
Busse, A. C., 257
Buswell, G. T., 211, 299, 325
Buswell-John Diagnostic Tests, 84, 211, 238

Central Tendency, Measures of, 127
Chance factor in objective tests, 111–113
Chapman, J. C., 114
Chapman-Cook Speed of Reading Tests, samples from, 26–28
Chemistry tests, 350
Clapp's Arithmetical Drills, 240
Clark, M. G., 284
Class diagnosis, 50, 60
Classification of pupils, 54–58
Class-intervals, 121–124
Cleveland Survey Tests, Arithmetic, 5, 26
Coefficient of Correlation, meaning of, 161–164
Compass Arithmetic Tests, diagnostic, 42–43, 191, 192, 193, 209, 210, 219–221; survey, 33–35
Composition, measurement of, 254
Correlation, computation of Pearson Product-Moment, 156–161; coefficient, meaning of, 161–164; meaning of, 152–155; table, 155
Courtis, S. A., 65, 250; arithmetic tests, 26, 209; practice tests in arithmetic, 240; reading tests, 26

Crabbs, Leah M., 322
Cram, Fred D., 355

Dearborn, W. F., 4, 8, 299
Deciles, 169–170
Derived scores, 194–205
Detroit First Grade Intelligence Tests, 23
Deviation, *See variability*
DeVoss, J. C., 46, 206
Dewey, John, 218, 230
Diagnosis, function of, 209–211; meaning of, 208
Diagnostic tests, 39
Difficulty of exercises, 98
Directions for objective tests, 95
Distributing agencies for tests, 335–336
Double-entry of correlation table, 155
Double true-false exercises, 110
Drill materials, general vs. special, 212; criteria for, 213–229

Economy in Testing, 83
Economy Problem Solving Exercises, 245–246
Economy Remedial Exercise Cards, 52–53, 216, 222, 226, 242–244
Educational age, 188
Educational tests, characteristics of, 11; criteria for, 72–84; significance of, 81
Educational quotient, (*E.Q.*), 64, 201
English tests, 341–342
Examinations, traditional, 89
Exercises, arrangement of, 99; classification of, 103–105; difficulty of, 98; form of statement, 97; objective types, 106–111; types of, 106–111

Farwell, H. W., 114
Fisher, Rev. George, 2
Fletcher, J. M., 284
Foreign language tests, 344–347
Fractions drill unit, 224–225
Franseen Diagnostic Tests in Language, 255
Franzen, R., 64, 203, 206
Freeman, F. N., 322
French tests, 345–346
Frequency table, 125; steps in making, 120, 126

Gates, A. I., 286, 300; reading tests, 298, 304–308
General science, 350
Geography, 322; tests, 342
Geometry tests, 347–348; drills, 354
German tests, 346
Germane, C. E. and E. G., 325
Gilliland, A. R., 46, 66, 206
Goodykoontz, Bess, 317, 326
Grade norms, 185–188; interpretation of, 191–194
Grading system, 100–103
Grammar tests, 341–342
Graphic record card, 57
Gray, C. T., 8, 289, 325
Gray, W. S., 4, 286, 293, 325; Oral reading tests, 295–298; 301–303
Greene, Charles E., 114
Greene, H. A., 8, 40, 53, 74, 88, 114, 173, 175, 206, 213, 230, 245, 250, 284, 312, 326
Gregory, C. A., 46, 66, 206
Grouping of test scores, 117

Haggerty, M. E., Intelligence Examination, 22; Reading Examinations, 48
Hahn-Lackey Geography Scales, 342
Handwriting, drills, 354; scales, 4, 342–343
Henmon Latin Tests, 29
Hillegas Composition Scale, 254
Hines, H. C., 66, 175, 206
History tests, 343
Horn, Ernest, 261, 316, 317, 321, 326, 355
Hosic, J. F., 255, 284
Hotz Algebra Scales, 29

Hoyt, F. S., 251
Huey, E. B., 299, 326
Hull, Clark L., 163

Illinois Examination, 22, 202
Index, ability to use, 314–315
Individual measurement, 20–21
Individual pupil diagnosis, 51–54, 62
Informal vs. standardized tests, 47
Intelligence Quotient (*I.Q.*), 63, 201
Intelligence tests, 344–345
Inter-dependence of mental and achievement tests, 63
Iowa High School Reading Tests, 74, 194, 312, 313, 314, 315
Iowa Placement Tests, 24
Iowa Spelling Scales, 15

John, Lenore, 211
Johnson, F. W., 8
Jordan, R. H., 46, 66, 206
Jorgensen, A. N., 74, 88, 312, 315
Journal of Educational Psychology, 88, 326
Journal of Educational Research, 175, 325–326
Judd, C. H., 326

Kelley, T. L., 32, 46, 66, 88, 206
Kelly, F. J., 8, 46, 206
Kirby Grammar Tests, 255
Klapper, Paul, 252, 253, 284, 326
Knight, F. B., 40, 53, 88, 114, 206, 230, 245, 250
Kuhlmann, F., 20
Kwalwasser Test of Music Information and Appreciation, 348
Kwalwasser-Ruch Musical Accomplishment Test, 348

Language, analysis of, 259–261; drill exercises, 276–283, 354; skills, 254–256; tests, 261–275
Latin tests, 29, 345
Lennes, N. J., 250
Lessenger, W. E., 203
Lippincott-Chapman Classroom Products Tests, 202

McBroom, Maude, 288, 326
McCall, W. A., 46, 64, 66, 88, 206
Mathematics aptitude, 24
Mathematics tests, 347
Mead, A. R., 88.
Mean, arithmetic, calculation of, 127–133
Measurable skills, in arithmetic, 235–237; in language, 259–261; in reading, 289–292
Measurement, pioneers in, 2
Median, calculation of, 135–137; definition of, 135; uses of, 137–138
Mental age, 198
Mental tests, 19, 22; uses of, 58–65
Meyer, Max, 9
Mid-measure, calculation of, 134; definition of, 134
Miles, W. R., 9
Modern language tests, 345–347
Monroe, W. S., 12, 46, 66, 88, 114, 175, 206; Silent Reading Tests, revised, 26, 187, 188
Morton, R. L., 250
Mueller, A. D., 287
Multiple-response exercises, 107–109
Music tests, 348

Nassau County Supplement to Hillegas Scale, 254
National Committee on Reorganization of English in Secondary Schools, 255
National Intelligence Tests, 344
New type tests, 89–114
Norms, 176; computation of, 180–183; kinds, 180; meaning of, 183; and standards, 176

Objectivity, 78–81, 93
Objective examinations, advantages of, 93–95; characteristics

of, 89–92; construction of, 95, 100; disadvantages of, 95; using results from, 100–103
Objective exercises, classification of, 103–105; samples of, 105–111
O'Brien, J. A., 287, 326
Odell, C. W., 175
Oral language, 255, 257–258
Oral reading, diagnosis of, 294–299; remedial drills, 299–303
Organization in language, 281–283
Otis, A. S., 21, 66, 175; Group Intelligence Tests, 22; Self-Administering Test, 22

Paterson, D. G., 105, 114
Pearson, Karl, 3; correlation method, 156–161
Percentile scores, 169–173; short method of computing, 171–173
Physics tests, 351
Power tests, 29
Pressey, L. W. and S. L., 66, 206
Pressey Classification Tests, 22; Diagnostic Tests in English Composition, 82
Preventive instruction, 211
Problem solving, 245–246; skills, 223; tests, 220–221
Prognostic tests, 23
Promotion and classification, 54–58

Q, as measure of variability, 141–144
Quality scores, 31
Quarterly Journal of Speech Education, 284
Quartiles, 142

Range, as measure of variability, 140; of scores, 120
Ranks, absolute, 169; assignment of, 165–173; relative, 166
Rapport, 21
Rate tests, 26
Reading, abilities, 288–293; defects, 293; diagnosis of, 294, 303–315; oral vs. silent, 294; tests, 348–350
Recall exercises, 106
Recognition type exercises, 107–111
Relationship, measures of, 152
Relative ranks, 166
Reliability, 76, 93
Remedial work, 209; in arithmetic, 231–250; language, 251–284; reading, 299–303
Remedial materials, criteria for, 213–229; in arithmetic, 240; 353–354; in language, 276–283; in reading, 299–303, 316–322, 355
Remmers, H. H., 115
Rice, J. M., 3
Rice, O. S., 317
Ruch, G. M., 12, 32, 40, 46, 53, 66, 88, 89, 92, 112, 115. 175, 206, 230, 235, 245, 25
Rugg, H. O., 175
Russell, Charles, 92, 206

Sampling, extent of, 94; illustration of, 77
Sangren-Woody Reading Test, 308–311
Scales, and tests, 14
School marks, 4–5
Science tests, 350–351
Scoring tests, 68, 86
Sears, J. B., 88
Seashore Music Talent Tests, 24
Sentences, remedial exercises in, 278
Shields, Grace, 316, 321, 326, 355
Sigma. *See Standard Deviation*
Silent Reading, skills, 288–293; measurement of, 303–313; remedial drills, 313–322
Smith, F. O., 9
Snedaker, Mabel I., 317, 326
Social science tests, 343
Spanish tests, 346–347

Spelling, remedial exercises, 276, 277; scales, Iowa, 15, 17–18; tests, 351
Standard Deviation, as measure of variability, 144–146; computation of, 146–152; definition of, 144
Standardization, 12, 180–183
Standards, vs. norms, 177–180; uses of, 179, 180
Stanford Achievement Tests, 32, 36–38, 202
Starch, Daniel, 9, 88
Statistical problems, 117, 356–374
Stenographic tests, 352
Stenquist Mechanical Aptitude Tests, 24
Stevenson Problem Analysis Tests, 219
Stoddard, G. D., 46, 66, 115, 175, 206
Studebaker, J. W., 40, 53, 88, 206, 230, 235, 245, 250
Stuttering, 258
Supervisory problems, 71–72
Surveys, 5
Survey tests, 32
Symonds, P. M., 175
Synthesis of skills, 227–228

Tabulation Methods, 117–127
Teacher's marks, 4–5
Teachers, and tests, 67
Terman, L. M., 20, 21, 32, 206; Group Test of Mental Ability, 22
Tests, characteristics of, 11; meaning of, 10–14, 328; and scales, 14; uses of, 48–65, 329–331
Testing and the Teacher, 67–69
Thorndike, E. L., 2, 4, 9, 21, 235, 250, 326
Thorndike-McCall Reading Tests, 15, 16, 29, 194, 196, 198
Thurstone, L. L., 175
Timing of tests, 86
Time standards for drills, 228–229
Toops, H. A., 88, 112
Trabue, M. R., 46, 66, 206
Travis, L. E., 257
True-False tests, 104, 109–111

Uhl, W. L., 326
Ullman-Kirby Latin Comprehension Tests, 345
Uses, of Achievement Tests, 48–58; of mental tests, 58–65

Validity, 73–76
Validation of tests, 73–76
Van Wagenen, W. J., 189
Variability, measures of, 139–151; Q as measure of, 141–144; range as measure of, 140; sigma as measure of, 144–151
Variation of scores, 116
Vocabulary tests, 298–299

West, P. V., 115
When to give tests, 69–70
Wiley, J. A., 326, 355
Willing Composition Scale, 15
Wilson Language Error Tests, 255
Wood, Ben D., 88, 92
Woody, Clifford, 206
Woody-McCall Mixed Fundamental Tests in Arithmetic, 30, 185–186, 206
Work-Book in Educational Measurements, 87, 127, 173, 193, 197
Wohlfarth, Julia H., 284
Writing scales, 342–343

Yerkes, R. M., 21
Yoakam, G. A., 326

DATE DUE

WITHDRAWN FROM
OHIO NORTHERN UNIVERSITY
LIBRARY

GAYLORD

PRINTED IN U.S.A.

HETERICK MEMORIAL LIBRARY
371.26 G79
onuu
Greene, Harry Andre/The use and interpre
3 5111 00057 2630

OCLC

Heterick Memorial Library
Ohio Northern University
Ada, Ohio 45810